THE ANNALS *of* MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

Vol. XX, No. 4 — December, 1949

THE ANNALS OF MATHEMATICAL STATISTICS

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore 2, Md. Subscriptions, renewals, orders for back numbers and other business communications should be sent to the ANNALS OF MATHEMATICAL STATISTICS, Mt. Royal & Guilford Aves., Baltimore 2, Md., or to the Secretary of the Institute of Mathematical Statistics, C. H. Fischer, Business Administration Building, University of Michigan, Ann Arbor, Mich.

Changes in mailing address which are to become effective for a given issue should be reported to the Secretary on or before the 15th of the month preceding the month of that issue. The months of issue are March, June, September and December.

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS after December 31, 1949 should be sent to T. W. Anderson, Department of Mathematical Statistics, Columbia University, New York 27, New York. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is $10.00 inside the Western Hemisphere and $5.00 elsewhere. Single copies $3.00. Back numbers are available at $10.00 per volume or $3.00 per single issue.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD., U. S. A.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the act of March 3, 1879.

LOCALLY BEST UNBIASED ESTIMATES[1]

By E. W. Barankin

University of California, Berkeley

Summary. The problem of unbiased estimation, restricted only by the postulate of section 2, is considered here. For a chosen number $s > 1$, an unbiased estimate of a function g on the parameter space, is said to be best at the parameter point θ_0 if its sth absolute central moment at θ_0 is finite and not greater than that for any other unbiased estimate. A necessary and sufficient condition is obtained for the existence of an unbiased estimate of g. When one exists, the best one is unique. A necessary and sufficient condition is given for the existence of only one unbiased estimate with finite sth absolute central moment. The sth absolute central moment at θ_0 of the best unbiased estimate (if it exists) is given explicitly in terms of only the function g and the probability densities. It is, to be more precise, specified as the l.u.b. of certain set $\mathfrak{A}$ of numbers. The best estimate is then constructed (as a limit of a sequence of functions) with the use of only the data (relating to g and the densities) associated with any particular sequence in $\mathfrak{A}$ which converges to the l.u.b. of $\mathfrak{A}$.

The case $s = \infty$ is considered apart. The case $s = 2$ is studied in greater detail. Previous results of several authors are discussed in the light of the present theory. Generalizations of some of these results are deduced. Some examples are given to illustrate the applications of the theory.

1. Introduction. Let Ω be a space of points x, and μ be a totally additive measure defined on a σ-field $\mathcal{F}$ of subsets of Ω. Let $\mathfrak{P} = \{p_\theta, \theta \in \Theta\}$ be a family of probability densities in Ω with respect to the measure μ. Θ is any index set; we lay down no conditions on its structure. We are concerned here with the existence and characterization of unbiased estimates of a real-valued function g on Θ, which are in some suitable sense "best" for a prescribed parameter point θ_0. That is, a real-valued, measurable (μ) function f_0 on Ω such that

$$\int_\Omega f_0 p_\theta \, d\mu = g(\theta), \qquad \theta \in \Theta, \tag{1}$$

and which satisfies a specified criterion of bestness for $\theta = \theta_0$. This criterion is usually taken to be

$$\int_\Omega (f_0 - g(\theta_0))^2 p_{\theta_0} \, d\mu \leqq \int_\Omega (f - g(\theta_0))^2 p_{\theta_0} \, d\mu, \qquad f \in \mathfrak{M}, \tag{2}$$

where $\mathfrak{M}$ denotes the class of all unbiased estimates of g; i.e., the class of all f satisfying (1). The obvious advantage in the definition (2) is the algebraic

[1] This article was prepared while the author was under contract with the Office of Naval Research.

pliability. The obvious disadvantage is that $\overline{\mathfrak{M}}$ may contain no estimate with finite variance (cf. section 9).

For the investigation of the fundamental questions, posed above, relating to unbiased estimates, we shall not restrict ourselves to (2). We consider chosen and fixed, a number $s > 1$, and lay down the

DEFINITION. *$f_0 \in \overline{\mathfrak{M}}$ is best at θ_0 if*

$$\infty > \int_\Omega |f_0 - g(\theta_0)|^s p_{\theta_0}\, d\mu \leqq \int_\Omega |f - g(\theta_0)|^s p_{\theta_0}\, d\mu, \qquad f \in \overline{\mathfrak{M}}.$$

With this, and under the condition of a rather natural postulate on $\mathfrak{P}$ (cf. section 2), we exhibit a necessary and sufficient condition for the existence of an unbiased estimate of g having a finite sth absolute central moment at θ_0.[2]

Except for the discussion, in section 3, of the case in which g is constant on Θ, we do not consider directly the estimation of g, but rather that of $h = g - g(\theta_0)$. Lemma 1, of section 2, gives the solution of the problem for g when that for h is known. After section 3, it is assumed exclusively that h is not $\equiv 0$, except where the contrary is explicitly stated.

In case s is finite, the existence theorem section 4, Theorem 2, asserts also the uniqueness of the best unbiased estimate of h. It is interesting to observe the similarity between the proof of this uniqueness and Fisher's proof of the (what might be called) asymptotic uniqueness of an efficient estimator [2 pp. 704, 705]. The case $s = \infty$[3] is discussed in section 5; in this case we find that, in general, the best estimate is not unique. However, for s both finite and infinite, and as well when g is constant ($\therefore\ h \equiv 0$), we give a necessary and sufficient condition that there be a unique unbiased estimate with finite s.a.c.m.[4] (cf. section 4, Corollary 2-1, and section 5, Theorem 3 (iii)).

Theorem 2 determines the s.a.c.m. of the best estimate as the l.u.b. of a set of numbers given explicitly; and thereby, in particular, throws open the class of all lower bounds of the minimum s.a.c.m. Investigations after such lower bounds, in the classical case $s = 2$, have led to the well-known results of Cramér-Rao [3 p. 480, (32.3.3)], and Bhattacharyya [4, p. 3, (1.10)]. In section 6, which is devoted to obtaining various special lower bounds, we show how those particular bounds fall out. It should be remarked, however, that our conditions on $\mathfrak{P}$ are in general different from those of the above authors.

---

[2] For the case $s = 2$ an alternative existence condition, antedating these results, but not yet published, has been obtained by C. Stein.

[3] If we use, in the above definition, the sth root of the sth absolute central moment, instead of the latter itself, then the bestness criterion for $s = \infty$ is the limiting criterion for $s \to \infty$; viz.,

$$\infty > \underset{x \in \Omega}{\text{ess. sup.}} |f_0 - g(\theta_0)| \leqq \underset{x \in \Omega}{\text{ess. sup}} |f - g(\theta_0)|, \qquad f \in \overline{\mathfrak{M}},$$

where ess. sup. refers to the measure $\nu(A) = \int_A p_{\theta_0}\, d\mu$.

[4] The abbreviation s.a.c.m. will henceforth be used to indicate sth absolute central moment at θ_0.

In section 7 we give, in Theorem 7 and its corollary, a construction of the best estimate, depending only on the knowledge of the minimum s.a.c.m. The latter, as indicated in the preceding paragraph, is always known independently of any knowledge of the best estimate. We use these results to obtain explicitly (Theorems 8 and 9) the best estimates, for arbitrary s, in two cases where we assume the minimum s.a.c.m. known. These cases, when $s = 2$, give the minimum variance as determined by the equality sign in the Cramér-Rao and Bhattacharyya inequalities, respectively.

Section 8 is given to a brief discussion of the special case $s = 2$. Finally, in section 9, we present a detailed study of an example.

At the suggestion of the referee we have added an appendix in which is given a brief running description of the fundamental ideas of Banach spaces that come into use here. The italicized phrases are those mentioned explicitly in the course of the paper.

We shall merely mention here certain points which will be elaborated further in future communications. (1) The general theory developed here pertains as well to sequential as to nonsequential estimation; one has only to make the proper identification of Ω, $\mathcal{F}$, μ, and $\mathfrak{P}$. Moreover, as applied to sequential estimation, the theory will determine the optimum stopping regions. (2) The discussion of section 5 below can be carried through with "ess. sup." referring to the measure μ, and $\mathfrak{L}_1$ being the space of functions on Ω which are integrable (μ); and for this, no restrictions whatsoever on the densities p_θ are required (cf. the postulate of section 2), since the p_θ are elements of this $\mathfrak{L}_1$ solely by virtue of their properties as probability densities. This development would, for example, be sufficient to yield the estimate of Girshick, Mosteller, and Savage [5] in the case of sequential binomial estimation. Also, this unrestricted analysis is fundamental for the problem of similar regions (a case of the bounded unbiased estimation of a constant function). (3) For any $s > 1$ it may be observed in the result of Theorem 7 below, that the best (at θ_0) estimate depends only on a sufficient statistic; this is clear from Neyman's theorem on sufficient statistics [6], since the best estimate depends only on ratios of the density functions p_θ. But more than this, Blackwell's method [7] of deriving a uniformly (over the parameter set) better unbiased estimate from a given unbiased estimate can be proved to remain valid also when the measure of dispersion is the sth absolute central moment, $s > 1$. And for this, the postulate of section 2 is not required. (4) Finally, we point out that, with the proper specializations of Θ, Cramér's theorem on the ellipsoid of concentration [8], Bhattacharyya's multidimensional inequality [9], and the extensions of the Rao, Cramér, and Bhattacharyya bounds to sequential estimation—as, for example, by Blackwell and Girshick [1], Wolfowitz [10], and Seth [20]—can be drawn from Theorem 4 below.

The inspiration for the mode of analysis in the following pages, and the major part of its substance, come from F. Riesz: his book [11 Ch. III] and the article [12] (in particular sections 8–11 thereof). In strictly mathematical terminology, Theorems 2 and 3 are given in [11] for the sequence-spaces ℓ_r; and

Theorem 2 in [12] for the spaces $\mathfrak{L}_r$ of functions on the real interval [0, 1] with Lebesgue integrable rth powers. The proofs are given there for the case of a denumerable set Θ; in [12] an indication is given of the extension to a non-denumerable Θ. Our proof of Theorems 2 and 3, however, follows that given by Banach [13, p. 74] for the case of denumerable Θ. It is based on two results, a theorem of Hahn-Banach [13, p. 55, Theorem 4], and the representation theorem (suitable for the general type of $\mathfrak{L}_r$ that we consider) for bounded linear functionals on $\mathfrak{L}_r$ [14, p. 338, Theorem 46]. The first of these, and the representation theorem for any $r > 1$, spring in fact from the same article [12, p. 475] of Riesz. In the case $r = 1$, the representation theorem is due originally to Steinhaus [15]; in the case $r = 2$, it was developed simultaneously in 1907 by Riesz [16] and Fréchet [17].

Riesz' proofs of the sufficiency of the condition in Theorem 2 proceed by constructing an explicit sequence of functions on Ω which converge strongly in $\mathfrak{L}_s$ to the (in the present statistical terminology) best estimate. Precisely, if in Theorem 7 below, we take, for each $n = 1, 2, \cdots$, the numbers $\alpha_1^n, \alpha_2^n, \cdots, \alpha_{k_n}^n$ so that the expression

$$\frac{\left| \sum_{i=1}^{k_n} \alpha_i^n h(\theta_i^n) \right|}{\left\| \sum_{i=1}^{k_n} \alpha_i^n \pi_{\theta_n^i} \right\|_r}$$

is maximum, then the assertion of this theorem is that of Riesz. However, Theorem 7 is established here without this strict requirement on the α_i^n. The dropping of this restriction was essential for the proofs of Theorems 8 and 9. The latter two theorems are, in fact, proved with the use of Corollary 7–1, which is an even stronger result than Theorem 7. This corollary falls out of the proof of Theorem 7 immediately, in consequence of our use of Lemma 2 for that proof. The lemma, moreover, eases the proof of Theorem 7 markedly, in doing away with the need for any differentiation.

2. Preliminary considerations. We begin then by introducing the absolutely continuous (with respect to μ) measure, defined on $\mathfrak{F}$,

$$\nu(A) = \int_A p_{\theta_0} \, d\mu, \qquad A \in \mathfrak{F}.$$

A function ϕ is summable (ν) over Ω if and only if $\phi \cdot p_{\theta_0}$ is summable (μ) over Ω; and we have

$$\int_\Omega \phi \, d\nu = \int_\Omega \phi \cdot p_{\theta_0} \, d\mu,$$

(cf. [18, pp. 36–38]). Assuming that each of the ratios

$$\pi_\theta(x) = \frac{p_\theta(x)}{p_{\theta_0}(x)}, \qquad \theta \in \Theta$$

is defined almost everywhere (μ) throughout Ω, it follows that f is an unbiased estimate of g if and only if

$$\int_\Omega f\pi_\theta \, d\nu = g(\theta), \qquad \theta \in \Theta. \tag{3}$$

We define

$$h(\theta) = g(\theta) - g(\theta_0).$$

Since

$$\int_\Omega \pi_\theta \, d\nu = 1, \qquad \theta \in \Theta,$$

it is clear from (3) that f is an unbiased estimate of g if and only if $f - g(\theta_0)$ is an unbiased estimate of h. Moreover, f is best, for g, at θ_0 if and only if $f - g(\theta_0)$ is best, for h, at θ_0.

Define

$$r = \frac{s}{s-1},$$

and let $\mathfrak{L}_r$ and $\mathfrak{L}_s$ be the spaces, normed in the usual way, of real-valued functions on Ω, with summable (ν) absolute rth and sth powers, respectively. We denote the respective norms by $\|\ \|_r$ and $\|\ \|_s$; that is, if $\xi \in \mathfrak{L}_r$ and $\eta \in \mathfrak{L}_s$,

$$\| \xi \|_r = \left(\int_\Omega | \xi |^r \, d\nu \right)^{1/r},$$

and

$$\| \eta \|_s = \left(\int_\Omega | \eta |^s \, d\nu \right)^{1/s}.$$

We note that these spaces, for $s < \infty$, are weakly compact (cf. [21]). This property will be used in the proof of Theorem 7. Also, we shall make explicit use of the representation theorem for linear functionals on $\mathfrak{L}_r$ [14, p. 338, Theorem 46].

The assumptions on $\mathfrak{P}$, or on $\mathfrak{P}_0 = \{\pi_\theta, \theta \in \Theta\}$, will now be the following.

POSTULATE: *The functions π_θ are defined almost everywhere (μ) in Ω, and are elements of $\mathfrak{L}_r$.*

The foregoing considerations combine to give the following equivalence.

LEMMA 1. *$\phi_0 + g(\theta_0)$ is an unbiased estimate of g, which is best at θ_0, if and only if (i) ϕ_0 satisfies the equations*

$$\int_\Omega \phi \cdot \pi_\theta \, d\nu = h(\theta), \qquad \theta \in \Theta, \tag{4}$$

and (ii) when ϕ is any other function satisfying (4), we have

$$\| \phi \|_s \geqq \| \phi_0 \|_s ;$$

that is, if and only if ϕ_0 *is an unbiased estimate of h with minimum (finite) norm in* $\mathfrak{L}_s$. *The s.a.c.m. of* $\phi_0 + g(\theta_0)$ *is precisely* $\| \phi_0 \|_s$.

Starting with section 4, we shall deal directly with the estimation of h.

3. The case of constant g. Throughout the remainder (section 4 et seq.) of this article, the function h is assumed, unless the contrary is explicitly stated, to be non-constant; that is, since $h(\theta_0) = 0$, not $\equiv 0$. We can, and shall in this section, obtain the results of the desired kind for the case of a constant function g, by a brief, direct attack.

Let $g(\theta) \equiv g_0$, a constant. Then of course $h(\theta) \equiv 0$. One unbiased estimate of g is immediately obvious, viz., $f_1(x) \equiv g_0$. The s.a.c.m. of f_1 is 0.

There will exist other[5] unbiased estimates of g with finite s.a.c.m. if and only if there exist non-null unbiased estimates, in $\mathfrak{L}_s$, of $0 \equiv h$. That is, by virtue of the isomorphism between $\mathfrak{L}_s$ and the space of linear functionals on $\mathfrak{L}_r$, there will exist an unbiased estimate of g with finite s.a.c.m., distinct from f_1, if and only if there exists a non-null functional on $\mathfrak{L}_r$ which vanishes on the elements of $\mathfrak{P}_0 = \{\pi_\theta, \theta \in \Theta\}$. And a necessary and sufficient condition that such a functional exist is that $\mathfrak{P}_0$ be not a fundamental set in $\mathfrak{L}_r$ [13, p. 58, Theorem 7].

Observe finally that, in any case, f_1 is the unique unbiased estimate of g with vanishing s.a.c.m.

We collect these results in the following statement.

THEOREM 1. *If* $g(\theta) \equiv g_0$, *a constant, then there is a unique best unbiased estimate of* g; *viz.,* $f_1(x) \equiv g_0$. *And the s.a.c.m. of* f_1 *is* 0.

A necessary and sufficient condition that there exist no other unbiased estimates of g having finite s.a.c.m. is that the set $\mathfrak{P}_0$ *be fundamental in* $\mathfrak{L}_r$.

As an illustration of the ideas of this section, consider the following example: Ω is the real interval [0, 1]; μ is Lebesgue measure; Θ is the set of non-negative integers; and

$$p_\theta(x) = (\theta + 1)x^\theta.$$

And take $\theta_0 = 0$. Then, ν is again Lebesgue measure, and $\pi_\theta = p_\theta$ for each θ. For definiteness, take $r = 2$ (the results in this case are the same for any $r \geqq 1$). It is well-known that the non-negative integer powers of x form a fundamental set in $\mathfrak{L}_2$ on a finite real interval. That is, if ξ is a function on [0, 1], such that $\int_0^1 \xi^2 \, dx < \infty$, and if $\epsilon > 0$, then there exist an integer n and coefficients b_0,

[5] That is, distinct from f_1 in the sense of $\mathfrak{L}_s$; or, equivalently, differing from f_1 on a set of positive (ν) measure. Whenever, in the sequel, an equation $\xi_1 = \xi_2$ appears, for two functions ξ_1 and ξ_2 in $\mathfrak{L}_r$ or $\mathfrak{L}_s$, equality almost everywhere (ν) in Ω will be understood. It is a consequence of our postulate that if two functions on Ω are equal almost everywhere (ν), they are equal almost everywhere (ν'), where ν' is anyone of the measures $\nu'(A) = \int_A p_{\theta'} \, d\mu$, $\theta' \in \Theta$.

$b_1, \cdots, b_n$ such that

$$\int_0^1 \left(\xi - \sum_{i=0}^{n} b_i x^i\right)^2 dx < \epsilon.$$

Hence, in this case an unbiased estimate with finite variance at $\theta = 0$ is unique (as well for a non-constant function g as for one which is constant over Θ; cf. section 4, Corollary 2-1).

4. The main theorem for non-constant h. We shall denote by $\mathfrak{M}_s$ the class (or the set in $\mathfrak{L}_s$) of all unbiased estimates of h that belong to $\mathfrak{L}_s$.

THEOREM 2. (i) *A necessary and sufficient condition that $\mathfrak{M}_s$ be non-empty is that there exist a constant C such that for every set of n functions $\pi_{\theta_1}, \pi_{\theta_2}, \cdots, \pi_{\theta_n}$, in $\mathfrak{P}_0$, and every set of n real numbers $a_1, a_2, \cdots, a_n$, we have, for every $n = 1, 2, \cdots$,*

$$\left| \sum_{i=1}^{n} a_i h(\theta_i) \right| \leqq C \left\| \sum_{i=1}^{n} a_i \pi_{\theta_i} \right\|_r . \tag{5}$$

(ii) *For every $\phi \in \mathfrak{M}_s$ we have $\| \phi \|_s \geqq C_0$, where C_0 is the g.l.b. of the set of admissible constants C in (5).*

(iii) *If $\mathfrak{M}_s$ is non-empty there is a unique $\phi_0 \in \mathfrak{M}_s$ with $\| \phi_0 \|_s = C_0$. Thus, ϕ_0 is the unique unbiased estimate of h which is best at θ_0.*

The non-constancy of h clearly implies $C_0 > 0$.

The necessity of condition (5) is immediate. Suppose $\phi \in \mathfrak{M}_s$, so that ϕ satisfies equations (4); then, for any $\theta_1, \theta_2, \cdots, \theta_n$, and any real numbers $a_1, a_2, \cdots, a_n$,

$$\sum_{i=1}^{n} a_i h(\theta_i) = \int_\Omega \phi \cdot \sum_{i=1}^{n} a_i \pi_{\theta_i} \cdot d\nu.$$

By the Hölder inequality it follows that

$$\left| \sum_{i=1}^{n} a_i h(\theta_i) \right| \leqq \|\phi\|_s \cdot \left\| \sum_{i=1}^{n} a_i \pi_{\theta_i} \right\|_r .$$

Hence (5) is satisfied with $C = \| \phi \|_s$.

Part (ii) of the theorem is hereby proved as well.

Suppose $\mathfrak{M}_s$ non-empty, and ϕ_0, ϕ_1 in $\mathfrak{M}_s$, such that $\| \phi_0 \|_s = \| \phi_1 \|_s = C_0$. Then $1/2\ (\phi_0 + \phi_1) \in \mathfrak{M}_s$ and therefore

$$1/2\ \| \phi_0 + \phi_1 \|_s \geqq C_0 .$$

But, by the Minkowski inequality,

$$1/2\ \| \phi_0 + \phi_1 \|_s \leqq 1/2\ (\| \phi_0 \|_s + \| \phi_1 \|_s) = C_0 ,$$

Hence

$$\| \phi_0 + \phi_1 \|_s = \| \phi_0 \|_s + \| \phi_1 \|_s .$$

This equality implies $\phi_1 = \alpha\, \phi_0$ for some positive α. But since the norms of ϕ_0 and ϕ_1 are equal (and $\neq 0$) α must be unity. Thus the uniqueness of ϕ_0 is proved.

It remains now to prove, assuming (5) satisfied, the existence of ϕ_0. Consider the functional F on $\mathfrak{P}_0$ defined by

$$F(\pi_\theta) = h(\theta).$$

The Hahn-Banach theorem alluded to in section 1 (viz., [13, p. 55, Theorem 4]) has precisely (5) as a necessary and sufficient condition for the existence of a linear functional G on $\mathfrak{L}_r$ satisfying

$$\begin{aligned} &\text{(a)}\ G(\pi_\theta) = h(\theta),\ \theta \in \Theta; \\ &\text{(b)}\ \| G \| \leqq C; \end{aligned}$$

where $\| G \|$ is the norm of G, i.e.,

$$\| G \| = \underset{\xi \in \mathfrak{L}_r}{\text{l.u.b.}} \frac{| G(\xi) |}{\| \xi \|_r}.$$

In particular, taking $C = C_0$, there is a linear functional G_0 on $\mathfrak{L}_r$ with

$$\begin{aligned} &\text{(a}'\text{)}\ G_0(\pi_\theta) = h(\theta),\ \theta \in \Theta \\ &\text{(b}'\text{)}\ \| G_0 \| \leqq C_0 . \end{aligned}$$

But, for an element $\sum_{i=1}^{n} a_i \pi_{\theta i}$ in the linear manifold $[\mathfrak{P}_0]$ spanned by the π_θ,

$$G_0 \left(\sum_i a_i \pi_{\theta_i}\right) = \sum_i a_i h(\theta_i),$$

so that

$$\| G_0 \| \geqq \underset{\xi \in [\mathfrak{P}_0]}{\text{l.u.b.}} \frac{| G_0(\xi) |}{\| \xi \|_r} = C_0 .$$

Hence (b′) is replaced by the precise statement

$$\text{(b}''\text{)}\ \| G_0 \| = C_0 .$$

Now the representation theorem for linear functionals on $\mathfrak{L}_r$ asserts the existence of $\phi_0 \in \mathfrak{L}_s$, such that

$$G_0(\xi) = \int_\Omega \phi_0 \cdot \xi \, d\nu,$$

and

$$\| \phi_0 \|_s = \| G_0 \| = C_0 .$$

This taken with (a′) establishes the existence of $\phi_0 \in \mathfrak{L}_s$ satisfying

$$\left\{ \begin{aligned} &\int_\Omega \theta_0 \pi_\theta \, d\nu = h(\theta), \qquad \theta \in \Theta \\ &\| \phi_0 \|_s = C_0 . \end{aligned} \right.$$

and this completes the proof of the theorem.

It is readily seen that $\mathfrak{M}_s$ will consist of more than just ϕ_0 if and only if there exists a non-null functional on $\mathfrak{L}_r$ which vanishes on $\mathfrak{P}_0$. Our discussion in section 3 therefore enables us to assert the following.

COROLLARY 2-1. $\mathfrak{M}_s$, *when it is non-empty, consists of* ϕ_0 *alone if and only if* $\mathfrak{P}_0$ *is fundamental in* $\mathfrak{L}_r$.

A word is in order concerning the following two consequences of the boundedness of the measure ν: (i) if $\mathfrak{P}_0 \subset \mathfrak{L}_r$, then also $\mathfrak{P}_0 \subset \mathfrak{L}_{r'}$ for every $r' < r$; (ii) if $\phi \in \mathfrak{L}_s$ then also $\phi \in \mathfrak{L}_{s'}$ for every $s' < s$. Otherwise stated: (i′) if $\mathfrak{P}_0$ satisfies the postulate of section 2 for the number r, it likewise satisfies this postulate for every (admissible) $r' < r$; (ii′) if $\mathfrak{M}_s$ is non-empty, then $\mathfrak{M}_{s'}$ is non-empty for every $s' < s$. Regarding (i′) we shall make only the obvious remark that although $\mathfrak{P}_0$ satisfies the postulate for every $r' < r$, there may be values of $r' < r$ such that no C for (5) exists; this will be exemplified in section 9. Where (ii′) is concerned, it is clear that the non-emptiness of $\mathfrak{M}_s$ will not necessarily imply that $\mathfrak{P}_0 \subset \mathfrak{L}_{s'/s'-1}$ for every $s' < s$, even though for every such s' $\mathfrak{M}_{s'}$ is non-empty. If for every $\phi \in \Theta$ other than θ_0 we have $\pi_\theta \notin \mathfrak{L}_{s'/s'-1}$, for some particular $s' < s$, then we may have the situation in which there are elements in $\mathfrak{M}_{s'}$ with norms arbitrarily close to 0. However, this cannot be the case if (a) for some θ other than θ_0, $\pi_\theta \in \mathfrak{L}_{s'/s'-1}$, and (b) h does not vanish identically on Θ', the set of those θ for which $\pi_\theta \in \mathfrak{L}_{s'/s'-1}$. For, when these two conditions are satisfied, Theorem 2 applies to h as defined on Θ'; consequently there is a positive lower bound for the s'—norms of the unbiased estimates of h over Θ'. And since every element of $\mathfrak{M}_s'$ is, in particular, an unbiased estimate of h over Θ', it follows that the norms of those elements are bounded below by a positive number.

5. The case $s = \infty$ $(r = 1)$. Let $\mathfrak{M}_\infty$ denote the class of essentially bounded (ν) unbiased estimates of h; and let bestness at θ_0 be defined with respect to the essential absolute suprema of the elements of this class. That is, the unbiased estimate ϕ_0, of h, is best at θ_0 if

$$\operatorname*{ess.\,sup.}_{x \in \Omega} |\phi_0(x)| < \infty,$$

and if, when ϕ is another unbiased estimate of h, we have

$$\operatorname*{ess.\,sup.}_{x \in \Omega} |\phi_0(x)| \leqq \operatorname*{ess.\,sup.}_{x \in \Omega} |\phi(x)|.$$

The fundamental postulate for the functions π_θ is, in this case, that $\mathfrak{P}_0 \subset \mathfrak{L}_1$.

Now, $\mathfrak{L}_\infty$, the space of essentially bounded, measurable (ν) functions on Ω, normed by ess. sup., is the space of linear functionals on $\mathfrak{L}_1$ [14, p. 338]. Examination of the proof of Theorem 2 will show that that proof goes through also in the present case in all but one detail: we cannot here in general prove the uniqueness of the best estimate. The proof of uniqueness breaks down since the equality

$$\text{ess. sup.}\ |\phi_0(x) + \phi_1(x)| = \text{ess. sup.}\ |\phi_0(x)| + \text{ess. sup.}\ |\phi_1(x)|$$

does not imply that ϕ_1 is a constant multiple of ϕ_0. Of course, if $\mathfrak{P}_0$ is fundamental in $\mathfrak{L}_1$, we have a fortiori the uniqueness of the best estimate.

The results for the case $s = \infty$ are then the following.

THEOREM 3. (i) *A necessary and sufficient condition that $\mathfrak{M}_\infty$ be non-empty is that there exist a constant C such that for every set of n functions $\pi_{\theta_1}, \pi_{\theta_2}, \cdots, \pi_{\theta_n}$, in $\mathfrak{P}_0$, and every set of n real numbers $a_1, a_2, \cdots, a_n$, we have, for every $n = 1, 2, \cdots$,*

$$\left| \sum_{i=1}^{n} a_i h(\theta_i) \right| \leqq C \left\| \sum_{i=1}^{n} a_i \pi_{\theta_i} \right\|_1 .$$

(ii) *For every $\phi \in \mathfrak{M}_\infty$ we have $\| \phi \|_\infty \geqq C_0$, where C_0 is the g.l.b. of the set of admissible constants C above.*

(iii) *When $\mathfrak{M}_\infty$ is non-empty, it contains elements with norm equal to C_0. These are the best* (at θ_0) *unbiased estimates of h. When $\mathfrak{P}_0$ is not fundamental in $\mathfrak{L}_1$, there need not exist a unique best estimate.*

We close this section with the remark that Theorem 1 remains valid, as it stands, in the case $s = \infty$.

6. Particular lower bounds for the minimum s.a.c.m. In order to stress their significance in the statistical context, we shall give the statements of this section with the help of the symbol $\sigma_s(\phi)$ for the sth root of the s.a.c.m. of the unbiased estimate ϕ, of h. We have of course, the relation

$$\sigma_s(\phi) = \| \phi \|_s .$$

Now, one of the most important aspects of Theorem 2 is that it presents us immediately with an explicit evaluation of the minimum $\sigma_s(\phi)$ for all $\phi \in \mathfrak{M}_s$. We state the formula in the form of a theorem.

THEOREM 4. *Let $\mathfrak{R}$ denote the set of all real numbers. Then,*

$$\underset{\phi \in \mathfrak{M}_s}{\text{g.l.b.}}\, \sigma_s(\phi) = \underset{\substack{\theta_1, \theta_2, \cdots, \theta_n \in \Theta \\ a_1, a_2, \cdots, a_n \in \mathfrak{R} \\ n = 1, 2, \cdots}}{\text{l.u.b.}} \frac{\left| \sum_{i=1}^{n} a_i h(\theta_i) \right|}{\left\| \sum_{i=1}^{n} a_i \pi_{\theta_i} \right\|_r}$$

For brevity, let us set

$$\underset{\phi \in \mathfrak{M}_s}{\text{g.l.b.}}\, \sigma_s(\phi) = \sigma_s^{\min}.$$

Since this theorem expresses $\sigma_s^{\min}$ as the l.u.b. of an explicit set of numbers, it is clear that *the class of all lower bounds of $\sigma_s^{\min}$ is thereby thrown open to us.* It follows that, when $s = r = 2$ and our hypotheses on $\mathfrak{P}$ are fulfilled, the classical lower bounds of Cramér-Rao [3, p. 480] and Bhattacharyya [4, p. 3] are particularized consequences of Theorem 4. In the results that follow here we shall indicate the deduction of those classical bounds. We need not, however, restrict s.

For a moment, let us denote by $\pi(x)$ the function on Θ which assigns the value $\pi_\rho(x)$ to the point $\rho \in \Theta$, and let Θ be an interval on the real axis. Then we shall, below, write π_θ' for the function (when it exists) on Ω which assigns the

value $(d\pi(x)/d\rho)_{\rho=\theta}$ to $x \epsilon \Omega$. Similarly, π_θ'' for the function assigning the value $(d^2\pi(x)/d\rho^2)_{\rho=\theta}$ to x; and so on.

THEOREM 5. *Suppose the following conditions fulfilled:*

(i) $\Theta = \mathcal{I}$, *an interval on the real axis;*

(ii) h *is differentiable on* $\Theta' \subseteqq \mathcal{I}$;

(iii) *for each* $\theta \epsilon \Theta'$, π_θ' *is defined almost everywhere* (ν), *and is an element of* $\mathfrak{L}_r$;

(iv) *for each* $\theta \epsilon \Theta'$,

$$\lim_{\rho \to \theta} \left\| \frac{\pi_\rho - \pi_\theta}{\rho - \theta} - \pi_\theta' \right\|_r = 0.$$

Then, for any $m + n$ $(m, n = 1, 2, \cdots)$ *points* $\theta_1, \theta_2, \cdots, \theta_m$ *in* $\mathcal{I}$, *and* $\theta_1', \theta_2', \cdots, \theta_n'$ *in* Θ', *and any* $m + n$ *real numbers* $a_1, a_2, \cdots, a_m, b_1, b_2, \cdots, b_n$ *such that*

$$\left\| \sum_{i=1}^m a_i \pi_{\theta_i} + \sum_{i=1}^n b_i \pi_{\theta_i'}' \right\|_r \neq 0,$$

we have

$$\sigma_s^{\min} \geqq \frac{\left| \sum_{i=1}^m a_i h(\theta_i) + \sum_{i=1}^n b_i h'(\theta_i') \right|}{\left\| \sum_{i=1}^m a_i \pi_{\theta_i} + \sum_{i=1}^n b_i \pi_{\theta_i'}' \right\|_r}. \tag{6}$$

The prime on the h in (6) denotes the derivative of h.

To prove this theorem, observe first that by virtue of Theorem 4, we may write

$$\sigma_s^{\min} \geqq \frac{\left| \sum_{i=1}^m a_i h(\theta_i) + \sum_{i=1}^n b_i \frac{h(\rho_i) - h(\theta_i')}{\rho_i - \theta_i'} \right|}{\left\| \sum_{i=1}^m a_i \pi_{\theta_i} + \sum_{i=1}^n b_i \frac{\pi_{\rho_i} - \pi_{\theta_i'}}{\rho_i - \theta_i'} \right\|_r}$$

for every set of points $\rho_1, \rho_2, \cdots, \rho_n$ in $\mathcal{I}$ such that the denominator of the right-hand side is defined and $\neq 0$. Therefore, also

$$\sigma_s^{\min} \geqq \lim_{\substack{\rho_i \to \theta_i \\ i=1,2,\cdots,n}} \frac{\left| \sum_{i=1}^m a_i h(\theta_i) + \sum_{i=1}^n b_i \frac{h(\rho_i) - h(\theta_i')}{\rho_i - \theta_i'} \right|}{\left\| \sum_{i=1}^m a_i \pi_{\theta_i} + \sum_{i=1}^n b_i \frac{\pi_{\rho_i} - \pi_{\theta_i'}}{\rho_i - \theta_i'} \right\|_r}. \tag{7}$$

Now, by condition (iv), the element

$$\sum_{i=1}^m a_i \pi_{\theta_i} + \sum_{i=1}^n b_i \frac{\pi_{\rho_i} - \pi_{\theta_i'}}{\rho_i - \theta_i'}$$

of $\mathfrak{L}_r$ converges, in the strong sense in $\mathfrak{L}_r$, to

$$\sum_{i=1}^m a_i \pi_{\theta_i} + \sum_{i=1}^n b_i \pi_{\theta_i'}',$$

as $\rho_i \to \theta_i'$, $i = 1, 2, \cdots, n$. Consequently we have convergence of the norm; that is, the denominator of the right-hand side of (7) converges to the denominator of (6). (The latter is $\neq 0$, so that for all ρ_i sufficiently close to θ_i', $i = 1, 2, \cdots, n$, the ratios in (7) are defined.) There is no difficulty about the convergence of the numerator of (7) to that of (6). The theorem is thus proved.

COROLLARY 5-1. *Under the hypothesis of Theorem* 5, *we have, in particular, when* $\theta_0 \in \Theta'$ *and* $\| \pi_{\theta_0}' \|_r \neq 0$,

$$\sigma_s^{\min} \geqq \frac{|h'(\theta_0)|}{\| \pi_{\theta_0}' \|_r}. \tag{8}$$

If we denote by p the function on $\Omega \times \Theta$ which assigns the value $p_\theta(x)$ to the point (x, θ), and write (8) in the form

$$(\sigma_s^{\min})^r \geqq \frac{|h'(\theta_0)|^r}{\displaystyle\int_\Omega \left| \frac{\partial \log p}{\partial \theta} \right|_{\theta=\theta_0}^r p_{\theta_0}\, d\mu}, \tag{8'}$$

the generalization of the Cramér-Rao inequality afforded by (8) becomes evident.

Using the result and method of Theorem 5, we can establish the next in a hierarchy of theorems.

THEOREM 6. *Suppose the hypothesis of Theorem* 5 *satisfied, and the following condition fulfilled: for each* θ *in a non-empty subset* Θ'' *of* Θ', (i) $h''(\theta)$ (*the second derivative*) *exists and* (ii) π_θ'' *is defined almost everywhere* (ν), *is an element of* $\mathfrak{L}_r$, *and satisfies*

$$\lim_{\rho \to \theta} \left\| \frac{\pi_\rho' - \pi_\theta'}{\rho - \theta} - \pi_\theta'' \right\|_r = 0.$$

Then, for any $m + n + q (m, n, q = 1, 2, \cdots)$ *points* $\theta_1, \theta_2, \cdots, \theta_m$ *in* $\mathcal{G}$, $\theta_1', \theta_2', \cdots, \theta_n'$ *in* Θ', *and* $\theta_1'', \theta_2'', \cdots, \theta_q''$ *in* Θ'', *and any* $m + n + q$ *real numbers* $a_1, a_2, \cdots, a_m, b_1, b_2, \cdots, b_n, c_1, c_2, \cdots, c_q$ *such that*

$$\left\| \sum_{i=1}^m a_i \pi_{\theta_i} + \sum_{i=1}^n b_i \pi_{\theta_i'}' + \sum_{i=1}^q c_i \pi_{\theta_i''}'' \right\|_r \neq 0,$$

we have

$$\sigma_s^{\min} \geqq \frac{\left| \sum_{i=1}^m a_i h(\theta_i) + \sum_{i=1}^n b_i h'(\theta_i') + \sum_{i=1}^q c_i h''(\theta_i'') \right|}{\left\| \sum_{i=1}^m a_i \pi_{\theta_i} + \sum_{i=1}^n b_i \pi_{\theta_i'}' + \sum_{i=1}^q c_i \pi_{\theta_i''}'' \right\|_r}.$$

Just as in the case of the previous theorem, we have here an immediate corollary.

COROLLARY 6-1. *Under the hypothesis of Theorem* 6, *we have in particular, when* $\theta_0 \in \Theta' \cdot \Theta''$,

$$\sigma_s^{\min} \geqq \frac{|bh'(\theta_0) + ch''(\theta_0)|}{\| b\pi_{\theta_0}' + c\pi_{\theta_0}'' \|_r}, \tag{9}$$

for any two real numbers, b and c, such that the denominator of the right-hand side does not vanish.

Consider (9) in the particular case $s = r = 2$. In this case, (9) may be written, explicitly,

$$(10) \qquad (\sigma_2^{\min})^2 \geqq \frac{|\,bh'(\theta_0) + ch''(\theta_0)\,|^2}{\displaystyle\int_\Omega \frac{1}{p_{\theta_0}}\left(b\,\frac{\partial p}{\partial \theta} + c\,\frac{\partial^2 p}{\partial \theta^2}\right)^2_{\theta=\theta_0} d\mu}.$$

In particular, (10) holds for values of b and c which maximize the right-hand side. And that maximum value is found, in the usual way, to be

$$J^{11}[h'(\theta_0)]^2 + 2J^{12}h'(\theta_0)h''(\theta_0) + J^{22}[h''(\theta_0)]^2,$$

where the matrix

$$\begin{pmatrix} J^{11} & J^{12} \\ J^{12} & J^{22} \end{pmatrix}$$

is the inverse of the matrix

$$\begin{pmatrix} \displaystyle\int_\Omega \frac{1}{p_\theta}\left(\frac{\partial p}{\partial \theta}\right)^2 d\mu & \displaystyle\int_\Omega \frac{1}{p_\theta}\frac{\partial p}{\partial \theta}\frac{\partial^2 p}{\partial \theta^2}\,d\mu \\ \displaystyle\int_\Omega \frac{1}{p_\theta}\frac{\partial p}{\partial \theta}\frac{\partial^2 p}{\partial \theta^2}\,d\mu & \displaystyle\int_\Omega \frac{1}{p_\theta}\left(\frac{\partial^2 p}{\partial \theta^2}\right)^2 d\mu \end{pmatrix}.$$

Thus, we have

$$(11) \qquad (\sigma_2^{\min})^2 \geqq J^{11}[h'(\theta_0)]^2 + 2J^{12}h'(\theta_0)h''(\theta_0) + J^{22}[h''(\theta_0)]^2.$$

This is seen to be Bhattacharyya's result for the case of derivatives up to second order.

It is obvious how we extend Theorem 6 to obtain a similar result involving the functions $\pi_\theta, \pi_\theta', \pi_\theta'', \cdots, \pi_\theta^{(n)}$, for any assigned n. And it is thereafter clear how, in the case $s = r = 2$, Bhattacharyya's general inequality may be deduced.

It is clear that we can proceed from Theorem 4, under suitable conditions, to lower bounds for $\sigma_s^{\min}$ which involve integrals of the functions $\pi(x)$ (and the corresponding integrals of h) as well as the derivatives of these functions.

In closing this section we note that all the above considerations apply equally to the case $s = \infty$.

7. Determination of the best estimate. We shall now prove the following theorem, which provides an explicit construction of the best (at θ_0) estimate of h. We repeat that s is now taken to be finite.

THEOREM 7. *Let $\mathfrak{M}_s$ be non-empty, and ϕ_0 be the best (at θ_0) unbiased estimate of h. Let $\{\theta_i^n, i = 1, 2, \cdots, k_n\}$, $n = 1, 2, \cdots$, be a sequence of (finite) sets of points of Θ, and $\{\alpha_i^n, i = 1, 2, \cdots, k_n\}$, $n = 1, 2, \cdots$, a sequence of sets of real numbers, such that*

$$\lim_{n\to\infty} \frac{\left|\sum_{i=1}^{k_n} \alpha_i^n h(\theta_i^n)\right|}{\left\|\sum_{i=1}^{k_n} \alpha_i^n \pi_{\theta_i^n}\right\|_r} = C_0 = \|\phi_0\|_s = \sigma_s^{\min}.$$

Then the functions ζ_n :

$$\zeta_n(x) = \frac{\sum_{i=1}^{k_n} \alpha_i^n h(\theta_i^n)}{\left\| \sum_{i=1}^{k_n} \alpha_i^n \pi_{\theta_i^n} \right\|_r^r} \cdot \left| \sum_{i=1}^{k_n} \alpha_i^n \pi_{\theta_i^n}(x) \right|^{r/s} \operatorname{sgn}\left(\sum_{i=1}^{k_n} \alpha_i^n \pi_{\theta_i^n}(x) \right)$$

(*are elements of* $\mathfrak{L}_s$ *and*) *converge strongly in* $\mathfrak{L}_s$ *to* ϕ_0 .
The strong convergence here means precisely that

$$\lim_{n\to\infty} \int_\Omega | \zeta_n - \phi_0 |^s \, d\nu = 0.$$

Clearly, we may, with no loss in generality, assume the numbers α_i^n to be such that

$$\left\| \sum_{i=1}^{k_n} \alpha_i^n \pi_{\theta_i^n} \right\|_r = 1, \; n = 1, 2, \cdots . \tag{12}$$

We shall suppose this to be the case throughout the proof. Then the essential property of the θ_i^n and the α_i^n is that

$$\lim_{n\to\infty} \left| \sum_{i=1}^{k_n} \alpha_i^n h(\theta_i^n) \right| = C_0 . \tag{13}$$

And in this normalized situation, the functions ζ_n will be given by

$$\zeta_n(x) = \sum_{i=1}^{k_n} \alpha_i^n h(\theta_i^n) \cdot \left| \sum_{i=1}^{k_n} \alpha_i^n \pi_{\theta_i^n}(x) \right|^{r/s} \operatorname{sgn}\left(\sum_{i=1}^{k_n} \alpha_i^n \pi_{\theta_i^n}(x) \right). \tag{14}$$

That these functions are elements of $\mathfrak{L}_s$ is easily seen; in fact,

$$\| \zeta_n \|_s = \left| \sum_{i=1}^{k_n} \alpha_i^n h(\theta_i^n) \right| .$$

The proof of this theorem will consist mainly in the application of the following two lemmas.

LEMMA 2. *Let* $0 \neq \eta \in \mathfrak{L}_s$, *and* $\{\xi_n , n = 1, 2, \cdots\}$ *be a sequence of functions in* $\mathfrak{L}_r$ *such that*

$$\text{(i)} \quad \| \xi_n \|_r = 1, \qquad n = 1, 2, \cdots$$

$$\text{(ii)} \quad \lim_{n\to\infty} \int_\Omega \xi_n \eta \, d\nu = \| \eta \|_s .$$

Then ξ_n *converges strongly in* $\mathfrak{L}_r$ *to the function*

$$\xi_0 = \frac{1}{\| \eta \|_s^{s/r}} | \eta |^{s/r} \operatorname{sgn} \eta.$$

Let us observe first that

$$\int_\Omega \xi_0 \eta \, d\nu = \| \eta \|_s \tag{15}$$

and

$$\| \xi_0 \| = 1.$$

Furthermore, ξ_0 is the unique element with norm $\leqq 1$ in $\mathfrak{L}_r$ having the property (15). For, if also,

$$\int_\Omega \xi_0' \eta \, d\nu = \| \eta \|_s, \qquad \| \xi_0' \|_r \leqq 1,$$

we then have

$$\int_\Omega \tfrac{1}{2}(\xi_0 + \xi_0') \cdot \eta \, d\nu = \| \eta \|_s;$$

and from this,

$$\tfrac{1}{2} \| \xi_0 + \xi_0' \|_r \| \eta \|_s \geqq \| \eta \|_s .$$

That is,

$$\| \xi_0 + \xi_0' \|_r \geqq 2 \geqq \| \xi_0 \|_r + \| \xi_0' \|_r .$$

From this, and (Minkowski)

$$\| \xi_0 + \xi_0' \|_r \leqq \| \xi_0 \|_r + \| \xi_0' \|_r ,$$

we have

$$\| \xi_0 + \xi_0' \|_r = \| \xi_0 \|_r + \| \xi_0' \|_r .$$

Therefore, for some $a > 0$, $\xi_0' = a\xi_0$. But we must have $a = 1$ if ξ_0 and ξ_0' are both to satisfy (15), as assumed. Hence $\xi_0' = \xi_0$.

Now consider the sequence $\{\xi_n\}$. Choose a sub-sequence $\{\xi_{n_i}\}$ that converges weakly to, say, ξ'. Then $\| \xi' \|_r \leqq 1$. We have

$$\int_\Omega \xi' \eta \, d\nu = \lim_{i \to \infty} \int_\Omega \xi_{n_i} \eta \, d\nu = \| \eta \|_s .$$

Hence, $\xi' = \xi_0$. And since $1 = \| \xi_{n_i} \|_r \to 1 = \| \xi_0 \|_r$, it follows that ξ_{n_i} converges strongly to ξ_0 (cf. [13, p. 139, section 3]).

Suppose there is a subsequence $\{\xi_{n_i}'\}$ of $\{\xi_n\}$ such that

$$\| \xi_{n_i}' - \xi_0 \| > \delta > 0, \qquad i = 1, 2, \cdots .$$

We have, nonetheless, for this subsequence, the hypotheses of our lemma satisfied. We can therefore apply the argument of the previous paragraph to extract a subsequence of $\{\xi_{n_i}'\}$, which converges strongly to ξ_0. This is in obvious contradiction to the above δ-assumption, and the lemma is hereby proved.

LEMMA 3. *Lemma 2 remains true with the roles of $\mathfrak{L}_r$ and $\mathfrak{L}_s$ interchanged.*

This is obvious.

Returning now to the proof of Theorem 7, let us first, for the sake of brevity,

introduce the notation:

$$c_n = \sum_{i=1}^{k_n} \alpha_i^n\, h(\theta_i^n),$$

$$\gamma_n = \operatorname{sgn}\left(\sum_{i=1}^{k_n} \alpha_i^n\, h(\theta_i^n)\right),$$

$$\psi_n = \sum_{i=1}^{k_n} \alpha_i^n\, \pi_{\theta_i^n}.$$

From

$$\int_\Omega \phi_0\, \pi_\theta\, d\nu = h(\theta), \qquad \theta \in \Theta.$$

we easily obtain

$$\int_\Omega \phi_0\, \psi_n\, d\nu = c_n, \qquad n = 1, 2, \cdots,$$

which we may write

$$\int_\Omega \phi_0 \cdot \gamma_n\, \psi_n\, d\nu = |\, c_n \,|, \qquad n = 1, 2, \cdots.$$

Since $|\, c_n \,| \to \|\, \phi_0 \,\|_s$ (cf. (13)) and $\|\, \gamma_n \psi_n \,\|_r = 1$, $n = 1, 2, \cdots$, (cf. (12)), we have, by Lemma 2, that $\gamma_n \psi_n$ converges strongly to

$$\psi_0 = \frac{1}{C_0^{s/r}}\, |\, \phi_0 \,|^{s/r} \operatorname{sgn} \phi_0. \tag{16}$$

The functions (cf. (14))

$$\zeta_n = c_n\, |\, \psi_n \,|^{r/s} \operatorname{sgn} \psi_n$$

obviously satisfy

$$\int_\Omega \zeta_n \cdot \gamma_n\, \psi_n\, d\nu = |\, c_n \,|, \qquad n = 1, 2, \cdots$$

And from this we conclude that

$$\lim_{n \to \infty} \int_\Omega \zeta_n\, \psi_0\, d\nu = C_0,$$

or

$$\lim_{n \to \infty} \int_\Omega \frac{\zeta_n}{|\, c_n \,|}\, \psi_0\, d\nu = 1 = \|\, \psi_0 \,\|_r.$$

We may apply Lemma 3 to this result, since $\|\, \zeta_n / |\, c_n \,| \,\|_r = 1$, $n = 1, 2, \cdots$. And we thereby conclude that $\zeta_n / |\, c_n \,|$ converges strongly to

$$|\, \psi_0 \,|^{r/s} \operatorname{sgn} \psi_0,$$

which, on substituting from the definition (16) of ψ_0, we find to be just

$$\frac{\phi_0}{C_0}.$$

Since $| c_n | \to C_0$, it follows immediately that ζ_n converges strongly to ϕ_0, and the theorem is proved.

The following corollary is actually of greater use in applications than Theorem 7 itself, for the reason that it leaves no doubt about the form of $\lim \zeta_n$ (i.e., ϕ_0) when we know explicitly the form of $\lim \gamma_n \psi_n$.

COROLLARY 7–1. *Assume the hypothesis of Theorem 7. Then the functions*

$$\frac{\operatorname{sgn}\left(\sum_{i=1}^{k_n} \alpha_i^n\, h(\theta_i^n)\right)}{\left\| \sum_{i=1}^{k_n} \alpha_i^n \pi_{\theta_i^n} \right\|_r} \cdot \sum_{1=1}^{k_n} \alpha_i^n\, \pi_{\theta_i^n}$$

converge strongly, in $\mathfrak{L}_r$, to a function ψ_0, and

$$\phi_0 = C_0 \,| \psi_0 |^{r/s} \operatorname{sgn} \psi_0 .$$

This is clear from the proof of the theorem.

By way of illustrating the application of these results, we shall prove the following theorem.

THEOREM 8. *Assume the hypothesis of Theorem 5. And, further, let the equality sign hold in* (8). *Then,*

$$\phi_0(x) = \frac{h'(\theta_0)}{\| \pi'_{\theta_0} \|_r^r} \cdot | \pi'_{\theta_0}(x) |^{r/s} \operatorname{sgn} \pi'_{\theta_0}(x).$$

Since (8) is an equality, we may under the hypothesis of Theorem 5, consider that we have

$$C_0 = \lim_{n \to \infty} \frac{\left| \dfrac{1}{\rho_n - \theta_0} h(\rho_n) - \dfrac{1}{\rho_n - \theta_0} h(\theta_0) \right|}{\left\| \dfrac{1}{\rho_n - \theta_0} \pi_{\rho_n} - \dfrac{1}{\rho_n - \theta_0} \pi_{\theta_0} \right\|_r}. \tag{17}$$

where $\{\rho_n\}$ is a sequence in $\mathcal{J}$ converging to θ_0. The numerator of the right-hand side of (17), sans the vertical bars, converges to $h'(\theta_0)$ (which is $\neq 0$, since $C_0 \neq 0$); hence, for all sufficiently large n, that expression has the signum of $h'(\theta_0)$. The functions whose norms appear in the denominator of (17) we know to converge strongly in $\mathfrak{L}_r$ to π'_{θ_0} (by the hypothesis of Theorem 5). Hence, for this case, the function ψ_0 of Corollary 7-1 is

$$\psi_0 = \frac{\operatorname{sgn} h'(\theta_0)}{\| \pi'_{\theta_0} \|_r} \pi'_{\theta_0}.$$

Therefore, by the same corollary,

$$\phi_0(x) = \frac{| h'(\theta_0) |}{\| \pi'_{\theta_0} \|_r} \cdot \left| \frac{\operatorname{sgn} h'(\theta_0)}{\| \pi'_{\theta_0} \|_r} \pi'_{\theta_0} \right|^{r/s}$$

$$\cdot \operatorname{sgn} h'(\theta_0) \cdot \operatorname{sgn} \pi_{\theta_0}(x)$$

$$= \frac{h'(\theta_0)}{\| \pi'_{\theta_0} \|_r^r} \, | \pi'_{\theta_0}(x) |^{r/s} \operatorname{sgn} \pi'_{\theta_0}(x).$$

And this is the result asserted in the theorem.

The reader will have no difficulty in establishing, in the exact pattern of the preceding proof, the following.

THEOREM 9. *Assume the hypothesis of Theorem* 6. *And, further, let the equality sign hold in* (9) *for* $b = b_0$, $c = c_0$. [6] *Then,*

$$\phi_0(x) = \frac{b_0 h'(\theta_0) + c_0 h''(\theta_0)}{\| b_0 \pi'_{\theta_0} + c_0 \pi''_{\theta_0} \|_r^r} \cdot | b_0 \pi'_{\theta_0}(x) + c_0 \pi''_{\theta_0}(x) |^{r/s} \cdot \operatorname{sgn} (b_0 \pi'_{\theta_0}(x) + c_0 \pi''_{\theta_0}(x)).$$

It is evident that results of the type in these theorems may be built up as well with integrals over the parameter space.

A question of considerable practical importance is that of the rapidity of convergence of the ζ_n to ϕ_0. An answer to this question, on the level of generality we are maintaining in this study, consists in relating this convergence to that of the $| c_n |$ to C_0. In the case $s = r = 2$, the answer is immediate and exact:

$$\begin{aligned}
\| \zeta_n - \phi_0 \|_2^2 &= \int_\Omega (\zeta_n - \phi_0)^2 \, d\nu \\
&= \int_\Omega \zeta_n^2 \, d\nu - 2 \int_\Omega \phi_0 \zeta_n \, d\nu + \int_\Omega \phi_0^2 \, d\nu \\
&= | c_n |^2 - 2 | c_n |^2 + C_0^2 \\
&= C_0^2 - | c_n |^2.
\end{aligned}$$

Thus, if one unbiased estimate is known, it provides, since its norm is $\geqq C_0$, an upper bound for $\| \zeta_n - \phi_0 \|_2$. The same is true in the general case (any s) once we have established an upper bound, depending on C_0 and $| c_n |$, for $\| \zeta_n - \phi_0 \|_s$. But in the general case, a good upper bound does not seem to be so close at hand. There are indications of the direction in which one must proceed, and we hope to draw some significant results out of these before long.

8. The case $s = r = 2$. The particular aspects of this case (where bestness of an estimate has reference to its *variance*), which arise out of the coincidence of $\mathfrak{L}_r$ and $\mathfrak{L}_s$, merit some discussion. We shall denote the inner product, $\int_\Omega \xi \eta \, d\nu$, of two functions ξ and η in $\mathfrak{L}_2$, as usual by (ξ, η). Let $\{\mathfrak{P}_0\}$ denote the closed linear manifold in $\mathfrak{L}_2$ spanned by the π_θ.

THEOREM 10. *Let* $\mathfrak{M}_2$ *be non-empty. Then* ϕ_0 *is the unique element of* $\mathfrak{M}_2$ *which lies in* $\{\mathfrak{P}_0\}$.

[6] In the case $s = 2$, b_0 and c_0 are the values which render (11) an equality.

To begin with it is clear that the functions ζ_n of Theorem 7, in the present case $s = r = 2$, are all elements of $[\mathfrak{P}_0]$, the linear manifold spanned by the π_θ. Hence, since ϕ_0 is the strong limit of these elements, $\phi_0 \in \{\mathfrak{P}_0\}$.

Now suppose also $\phi_1 \in \mathfrak{M}_2$, $\phi_1 \in \{\mathfrak{P}_0\}$. Then, from

$$(\phi_0, \pi_\theta) = h(\theta), \qquad \theta \in \Theta,$$

$$(\phi_1, \pi_\theta) = h(\theta), \qquad \theta \in \Theta,$$

we have

$$(\phi_1 - \phi_0, \pi_\theta) = 0, \qquad \theta \in \Theta,$$

and, by continuity of the inner product,

$$(\phi_1 - \phi_0, \xi) = 0, \qquad \xi \in \{\mathfrak{P}_0\};$$

that is, $\phi_1 - \phi_0 \in \{\mathfrak{P}_0\}^\perp$. But, from $\phi_0 \in \{\mathfrak{P}_0\}$ and $\phi_1 \in \{\mathfrak{P}_0\}$ it follows that $\phi_1 - \phi_0 \in \{\mathfrak{P}_0\}$. Hence $\phi_1 - \phi_0 = 0$, and this proves the exclusiveness of the property for ϕ_0.

Another characterization of ϕ_0 is given by the following corollary.

Corollary 10-1. *If $\mathfrak{M}_2$ is non-empty, then ϕ_0 is the unique element of $\mathfrak{M}_2$ which satisfies the system of equations in ξ : $(\phi, \xi) = \|\xi\|_2^2$, $\phi \in \mathfrak{M}_2$.*

To see that ϕ_0 has the asserted property, let ϕ be any element of $\mathfrak{M}_2$, and set $\phi = \xi + \eta$, with $\xi \in \{\mathfrak{P}_0\}$ and $\eta \in \{\mathfrak{P}_0\}^\perp$. From

$$(\xi, \pi_\theta) = (\xi + \eta, \pi_\theta) = (\phi, \pi_\theta) = h(\theta),$$

it follows that $\xi \in \mathfrak{M}_2$. Hence $\xi = \phi_0$. And so,

$$(\phi, \phi_0) = (\phi_0 + \eta, \phi_0) = \|\phi_0\|_2^2.$$

If $\phi_1 \in \mathfrak{M}_2$ has this property also, then both

$$(\phi_1, \phi_0) = \|\phi_0\|_2^2$$

and

$$(\phi_0, \phi_1) = \|\phi_1\|_2^2;$$

and therefore

$$\|\phi_1\|_2 = \|\phi_0\|_2.$$

This proves $\phi_1 = \phi_0$, and so the corollary.

9. An example. Let Ω be Euclidean n-space, $x = (x_1, x_2, \cdots, x_n)$; μ, Lebesgue measure; Θ, the set of real numbers; and

$$p_\theta(x) = \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} (x_i - \theta)^2\right\}.$$

And finally, let $\theta_0 = 0$. Then

$$\pi_\theta(x) = \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} (-2\theta x_i + \theta^2)\right\}.$$

If $0 < b < \frac{1}{2}$, and we define

$$\phi_1(x) = (1 - 2b)^{n/2} \exp\left\{b \sum_{i=1}^{n} x_i^2\right\} - 1,$$

we have, for each θ,

$$\int_\Omega \phi_1(x) p_\theta(x) \, d\mu = \exp\left\{\frac{nb}{1 - 2b} \theta^2\right\} - 1.$$

Thus, ϕ_1 is an unbiased estimate of the function h:

$$h(\theta) = \exp\left\{\frac{nb}{1 - 2b} \theta^2\right\} - 1.$$

If we examine

$$\| \phi_1 \|_s^s = \frac{1}{(2\pi)^{n/2}} \int_\Omega \left| (1 - 2b)^{n/2} \exp\left\{b \sum_{i=1}^{n} x_i^2\right\} - 1 \right|^s \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} x_i^2\right\} d\mu;$$

we find that this integral converges only for $s < 1/2b$. Shifting the emphasis, we may state: *for the function h, defined by*

$$h(\theta) = e^{\alpha\theta^2} - 1, \qquad \alpha > 0,$$

there exists an unbiased estimate with finite sth *moment at $\theta = 0$, for each*

$$s < \frac{n + 2\alpha}{2\alpha}.$$

Next, observe that

$$\begin{aligned} \| \pi_\theta \|_r^r &= \frac{1}{(2\pi)^{n/2}} \int_\Omega \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} (x_i^2 - 2r\theta x_i + r\theta^2)\right\} d\mu \\ &= \exp\left\{\tfrac{1}{2} nr(r - 1)\theta^2\right\}, \end{aligned}$$

so that the π_θ are elements of $\mathfrak{L}_r$ for each $r > 1$. The ratio

$$\frac{| h(\theta) |}{\| \pi_\theta \|_r} = (e^{\alpha\theta^2} - 1) \exp\left\{-\tfrac{1}{2}n(r - 1)\theta^2\right\}$$

is seen to diverge as $\theta \to \infty$, if

$$\tfrac{1}{2} n(r - 1) < \alpha.$$

Hence, by Theorem 2, there exists no unbiased estimate of h belonging to $\mathfrak{L}_s$ for a value of s such that the number

$$r = \frac{s}{s - 1}$$

satisfies the inequality just above; that is, for a value of s greater than

$$\frac{n + 2\alpha}{2\alpha}.$$

Otherwise stated: *there exists no unbiased estimate of h with finite* sth *moment at* $\theta = 0$, *for*

$$s > \frac{n + 2\alpha}{2\alpha}.$$

It is most likely true that this last statement holds, in general, with

$$s \geqq \frac{n + 2\alpha}{2\alpha}.$$

We shall consider here only the case

$$\frac{n + 2\alpha}{2\alpha} = 2;$$

and since the analysis is the same for every pair n, α satisfying this equality, we treat the particular case of

$$n = 1, \qquad \alpha = \tfrac{1}{2}.$$

Thus, we shall show: *for* $n = 1$, *there exists no unbiased estimate of* h_2,

$$h_2(\theta) = e^{\frac{1}{2}\theta^2} - 1,$$

with finite variance at $\theta = 0$.

We must show that the ratios

$$\frac{\left| \sum_{i=1}^{m} a_i (e^{\frac{1}{2}\theta_i^2} - 1) \right|}{\left\| \sum_{i=1}^{m} a_i \pi_{\theta_i} \right\|_2}$$

are not bounded for all choices of m (distinct) θ_i's, and all sets of m real numbers a_i, and all m. This is clearly equivalent to showing the same for the ratios

$$Q(m, a_i, \theta_i) = \frac{\left| \sum_{i=1}^{m} a_i (1 - e^{-\frac{1}{2}\theta_i^2}) \right|}{\left\| \sum_{i=1}^{m} a_i e^{-\frac{1}{2}\theta_i^2} \pi_{\theta_i} \right\|_2}.$$

Now we find, by direct computation,

$$\left\| \sum_{i=1}^{m} a_i e^{-\frac{1}{2}\theta_i^2} \pi_{\theta_i} \right\|_2^2 = \sum_{i,j=1}^{m} e^{-\frac{1}{2}(\theta_i - \theta_j)^2} a_i a_j.$$

And the solution of the familiar extremum problem:

$$\sup_{(a_i)} \left| \sum_{i=1}^{m} a_i (1 - e^{-\frac{1}{2}\theta_i^2}) \right| \quad \text{subject to} \quad \sum_{i,j=1}^{m} e^{-\frac{1}{2}(\theta_i - \theta_j)^2} a_i a_j = 1$$

yields

$$\sup_{(a_i)} Q^2(m, a_i, \theta_i) = \sum_{i,j=1}^{m} v_{ij} (1 - e^{-\frac{1}{2}\theta_i^2})(1 - e^{-\frac{1}{2}\theta_j^2}),$$

where the matrix

$$V = (v_{ij}), \qquad i, j = 1, 2, \cdots, m,$$

is the inverse of the matrix

$$U = (e^{-\frac{1}{2}(\theta_i - \theta_j)^2}), \qquad i, j = 1, 2, \cdots, m.$$

We now take, in particular,

$$\theta_i = it, \qquad i = 1, 2, \cdots, m,$$

where t is a positive number. Clearly, there exists a number t_0 such that for $t > t_0$,

$$U(t) = (e^{-\frac{1}{2}(i-j)^2 t^2})$$

is non-singular. Also,

$$\lim_{t \to \infty} U(t) = I,$$

the identity matrix. Then, for $t > t_0$, $V = U^{-1}$ is a continuous function of U, so that

$$\lim_{t \to \infty} V(t) = (\lim_{t \to \infty} U(t))^{-1} = I.$$

Hence,

$$\lim_{t \to \infty} v_{ij}(t) = \delta_{ij}.$$

It follows that

$$\lim_{t \to \infty} \sup_{(a_i)} Q^2(m, a_i, it) = m,$$

and therefore,

$$\sup_{(a_i, \theta_i)} Q^2(m, a_i, \theta_i) \geqq m.$$

(A simple argument on the characteristic values of U shows that there is actually equality here.) This result gives the unboundedness of the ratios Q_j and our proposition is proved, by virtue of Theorem 2.

APPENDIX

The spaces $\mathfrak{L}_r$ and $\mathfrak{L}_s$ are instances of a Banach space over the reals; that is, a complete, normed, linear vector space, closed under multiplication by real numbers. That the space, say $\mathfrak{B}$, is normed is to say that there is a non-negative, real-valued function, $\|\ \|$, defined on $\mathfrak{B}$, with the properties:

$$\| \xi \| = 0 \quad \text{if and only if } \xi \text{ is the null vector},$$

$$\| a\xi \| = | a | \cdot \| \xi \|,$$

$$\| \xi + \eta \| \leqq \| \xi \| + \| \eta \|;$$

where ξ, $\eta \in \mathfrak{B}$ and a is real. The number $\| \xi \|$ is called the *norm* of ξ.

The function $\| \xi - \eta \|$ on pairs ξ, η of vectors is a distance function in the usual sense. With it, *strong convergence* (or simply *convergence*) is defined in $\mathfrak{B}$: ξ_n *converges strongly* to ξ when $\lim_{n\to\infty} \| \xi_n - \xi \| = 0$. In symbols: $\xi_n \to \xi$ or $\lim \xi_n = \xi$. The usual set-theoretic notions are now defined in the obvious way; e.g., limit point of a set, closed set, etc. That the space $\mathfrak{B}$ is complete means that every sequence $\{\xi_n\}$ satisfying $\lim_{m,n\to\infty} \| \xi_m - \xi_n \| = 0$ converges to a (unique) element $\xi \in \mathfrak{B}$.

A *linear manifold* $\mathfrak{M}$ in $\mathfrak{B}$ is a subset of $\mathfrak{B}$ with the property that for any two elements ξ, $\eta \in \mathfrak{M}$ and any two real numbers a, b, we have also $a\xi + b\eta \in \mathfrak{M}$. A *closed linear manifold* is a linear manifold that is closed in the set-theoretic sense. If S is any subset of $\mathfrak{B}$, then the set, $[S]$, of all finite linear combinations of elements of S is a linear manifold; it is the *linear manifold spanned by* S. The closure of $[S]$, denoted by $\{S\}$, is called the *closed linear manifold spanned by* S. In general, $[S]$ is a proper subset of $\{S\}$. A set $S \subseteq \mathfrak{B}$ is called *fundamental* when $\{S\} = \mathfrak{B}$.

A *linear functional*, G, on $\mathfrak{B}$ is a real-valued function with the property that for any two elements ξ, $\eta \in \mathfrak{B}$ and any two real numbers a, b, we have $G(a\xi + b\eta) = aG(\xi) + bG(\eta)$. The linear functional G is said to be *bounded* when the number

$$\| G \| = \underset{\| \xi \| \neq 0}{\text{l.u.b.}} \frac{| G(\xi) |}{\| \xi \|}$$

is finite. $\| G \|$ is called the *norm* of G. (Throughout the text of the paper, the qualification "bounded" has been understood in all references to linear functionals). If we define the sum of two linear functionals F and G by $(F + G)(\xi) = F(\xi) + G(\xi)$, and make the other requisite definitions in the obvious way, we find that the bounded linear functionals on $\mathfrak{B}$ form a linear vector space over the reals. The function $\| \; \|$ on the bounded linear functionals, which we have already called a norm, is in fact a norm in the Banach space sense. This vector space, so normed, is readily shown to be complete. Hence it is a Banach space—usually called the conjugate space to $\mathfrak{B}$. It is this space we have referred to in the text as the *space of linear functionals on* $\mathfrak{B}$.

If a sequence $\{\xi_n\}$ of elements of $\mathfrak{B}$ has the property that $\lim_{n\to\infty} G(\xi_n) = G(\xi)$ for every bounded linear functional G, then ξ_n is said to *converge weakly to* ξ. If, of the sequence $\{\xi_n\}$, we know only that $\lim_{n\to\infty} G(\xi_n)$ exists for every bounded linear functional, we say simply that the sequence is weakly convergent. The space $\mathfrak{B}$ is called weakly complete if every weakly convergent sequence converges weakly to a limit. The spaces $\mathfrak{L}_r$, $r \geqq 1$ are weakly complete. $\mathfrak{B}$ is said to be *weakly compact* if every bounded set $S \subset \mathfrak{B}$ contains a weakly convergent sequence. That S is "bounded" means $\underset{\xi \in S}{\text{l.u.b.}} \| \xi \| < \infty$.

A real Hilbert space $\mathfrak{H}$ is a real Banach space on which there is defined an

inner product; that is, a function (ξ, η) on pairs of elements ξ, η, with the properties

$$(\xi, \eta) = (\eta, \xi),$$

$$(a\xi, \eta) = a(\xi, \eta),$$

$$(\xi + \zeta, \eta) = (\xi, \eta) + (\zeta, \eta),$$

$$\| \xi \|^2 = (\xi, \xi).$$

The inner product is a *continuous* function of both its arguments; i.e., $\lim \xi_m = \xi$ and $\lim \eta_n = \eta$ imply $\lim (\xi_m , \eta_n) = (\xi, \eta)$. The space $\mathfrak{L}_2$ in the text is a Hilbert space when we take $(\xi, \eta) = \int_\Omega \xi\eta \, d\nu$. Two elements ξ, η which are such that $(\xi, \eta) = 0$ are said to be orthogonal. If S is any set in $\mathfrak{H}$, then the set of elements of $\mathfrak{H}$ each of which is orthogonal to every element of S is called the *orthocomplement* of S, and is denoted by $S^\perp$.

For further elaboration the reader is referred to [13] and [19].

REFERENCES

[1] D. Blackwell and M. A. Girshick, "A lower bound for the variance of some unbiased sequential estimates," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 277–280.

[2] R. A. Fisher, "Theory of statistical estimation," *Camb. Phil. Soc. Proc.*, Vol. 22, (1925), pp. 700–725.

[3] H. Cramér, *Mathematical Methods of Statistics*, Princeton Press, Princeton, 1946.

[4] A. Bhattacharyya, "On some analogues of the amount of information and their use in statistical estimation," *Sankhyā*, Vol. 8 (1946), pp. 1–14.

[5] M. A. Girshick, F. Mosteller, and L. J. Savage, "Unbiased estimates for certain binomial sampling problems," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 13–23.

[6] J. Neyman, "Su un teorema concernente le cosidette statistiche sufficienti," *Giornale dell'Istituto Italiano degli Attuari*, Vol. 6 (1935), pp. 320–334.

[7] D. Blackwell, "Conditional expectation and unbiased sequential estimation," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 105–110.

[8] H. Cramér, "A contribution to the theory of statistical estimation," *Skand. Aktuar. tids.*, Vol. 29 (1946), pp. 85–94.

[9] A. Bhattacharyya, "On some analogues of the amount of information and their use in statistical estimation (cont'd)," *Sankhyā*, Vol. 8 (1947), pp. 201–218.

[10] J. Wolfowitz, "The efficiency of sequential estimates and Wald's equation for sequential processes," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 215–230.

[11] F. Riesz, *Les Systèmes d'Équations Linéaires a une Infinité d'Inconnues*, Gauthier-Villars, Paris, 1913.

[12] F. Riesz, "Untersuchungen über Systeme integrierbare Funktionen," *Math. Annalen*, Vol. 69 (1910), pp. 449–497.

[13] S. Banach, *Théorie des Opérations Linéaires*, Garasinski, Warsaw, 1932.

[14] N. Dunford, "Uniformity in linear spaces," *Am. Math. Soc. Trans.*, Vol. 44 (1938), pp. 305–356.

[15] M. H. Steinhaus, "Additive und stetige funktionaloperationen," *Math. Zeits.*, Vol. 5 (1918), pp. 186–221.

[16] F. Riesz, "Sur une espèce de géométrie analytique des systèmes de fonctions sommables," *Comptes Rendus*, Vol. 144 (1907), pp. 1409–1411.
[17] M. Fréchet, "Sur les ensembles de fonctions et les opérations linéaires," *Comptes Rendus*, Vol. 144 (1907), pp. 1414–1416.
[18] S. Saks, *Theory of the Integral*, Stechert, New York, 1937.
[19] J. von Neumann, *Functional Operators*, (Mimeographed notes) Princeton, 1935.
[20] G. R. Seth, "On the variance of estimates," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 1–27.
[21] B. J. Pettis, "A note on regular Banach spaces," *Am. Math. Soc. Bull.*, Vol. 44 (1938), pp. 420–428.

A SEQUENTIAL DECISION PROCEDURE FOR CHOOSING ONE OF THREE HYPOTHESES CONCERNING THE UNKNOWN MEAN OF A NORMAL DISTRIBUTION

By Milton Sobel and Abraham Wald[1]

Columbia University

1. Introduction. In this paper a multi-decision problem is investigated from a sequential viewpoint and compared with the best non-sequential procedure available. Multi-decision problems occur often in practice but methods to deal with such problems are not yet sufficiently developed.

The problem under consideration here is a 3-decision problem: Given a chance variable which is normally distributed with known variance σ^2, but unknown mean θ, and given two real numbers $a_1 < a_2$, the problem is to choose one of the three mutually exclusive and exhaustive hypotheses

$$H_1 : \theta < a_1 \qquad H_2 : a_1 \leqq \theta \leqq a_2 \qquad H_3 : \theta > a_2 .$$

In order to select a proper sequential decision procedure, the parameter space is subdivided into 5 mutually exclusive and exhaustive zones in the following manner. Around a_1 there exists an interval (θ_1, θ_2) in which we have no strong preference between H_1 and H_2 but prefer (strongly) to reject H_3. Around a_2 there exists an interval (θ_3, θ_4) in which we have no strong preference between H_2 or H_3 but prefer (strongly) to reject H_1. For $\theta \leqq \theta_1$ we prefer to accept H_1. For $\theta_2 \leqq \theta \leqq \theta_3$ we prefer to accept H_2. For $\theta \geqq \theta_4$ we prefer to accept H_3.

The intervals (θ_1, θ_2) and (θ_3, θ_4) will be called indifference zones. The determination of these indifference zones is not a statistical problem but should be made on practical considerations concerning the consequences of a wrong decision.

In accordance with the above we define a wrong decision in the following way. For $\theta \leqq \theta_1$, acceptance of H_2 or H_3 is wrong. For $\theta_1 < \theta < \theta_2$ acceptance of H_3 is wrong. For $\theta_2 \leqq \theta \leqq \theta_3$, acceptance of H_1 or H_3 is wrong. For $\theta_3 < \theta < \theta_4$, acceptance of H_1 is wrong. For $\theta \geqq \theta_4$, acceptance of H_1 or H_2 is wrong.

The requirements on our decision procedure necessary to limit the probability of a wrong decision are investigated. Two cases are considered.

Case 1: Prob. of a wrong decision $\leqq \gamma$ for all θ.

$$\text{Case 2: } \begin{cases} \text{Prob. of a wrong decision } \leqq \gamma_1 \text{ for } \theta \leqq \theta_1 , \\ \text{Prob. of a wrong decision } \leqq \gamma_2 \text{ for } \theta_1 < \theta < \theta_4 , \\ \text{Prob. of a wrong decision } \leqq \gamma_3 \text{ for } \theta \geqq \theta_4 . \end{cases}$$

The decision procedure discussed in the present paper is not an optimum procedure since, as will be seen later, the final decision at the termination of

[1] Work done under the sponsorship of the Office of Naval Research.

experimentation is not in every case a function of only "the sample mean of *all* the observations", although the sample mean is a sufficient statistic for θ. Although the procedure considered is not optimal it is suggested for the following reasons:

1. The decision procedure can be carried out simply. In fact tables can be constructed before experimentation starts that render the procedure completely mechanical.

2. The derivation of the operating characteristic (OC) function, neglecting the excess of the cumulative sum over the boundary, is accomplished with little difficulty. In general, for other multi-decision problems it is unknown how to obtain the OC function.

3. It is believed that the loss of efficiency is not serious; i.e., the suggested sequential procedure is not far from being optimum. In this connection a non-sequential procedure is compared with this sequential procedure. The results show that, for the same maximum probability of making a wrong decision, the sequential procedure requires on the average substantially fewer observations to reach a final decision. In fact, for Case 1 noted above, if $.008 < \gamma < .1$, and if certain symmetrical features are assumed, then the fixed number of observations required by the non-sequential method is greater than the maximum of the average sample number (ASN) function taken over all values of θ.

It was found necessary in the course of the investigation to put an upper bound on the quantity $\dfrac{\theta_4 - \theta_3}{a_2 - a_1}$ in order that the methods used to obtain upper and lower bounds for the ASN function should give close results. This restriction, however, is likely to be satisfied in practical applications.

All formulas for ASN and OC functions which will be used in this paper will be approximation formulas neglecting the excess of the cumulative sum over the boundaries. Nevertheless, equality signs will be used in these formulas, except when additional approximations are involved.

2. Description of the Decision Procedure.[2] We shall assume that the indifference zones described above have the following properties

$$\text{(i)} \quad \theta_1 < a_1 < \theta_2 \leqq \theta_3 < a_2 < \theta_4$$

$$\text{(ii)} \quad \theta_1 + \theta_2 = 2a_1 \; ; \quad \theta_3 + \theta_4 = 2a_2$$

$$\text{(iii)} \quad \theta_2 - \theta_1 = \theta_4 - \theta_3 = \Delta \text{ (say)}.$$

[2] A similar decision procedure was used by P. Armitage [2] as an alternative to the sequential t test (with 2-sided alternatives). The form used there is more restricted as he considers only the case $\theta_2 = \theta_3$. Essential inequalities on the OC function are pointed out but no attempt is made to determine the complete OC and ASN functions. A closely related but somewhat different procedure for dealing with a trichotomy was suggested by Milton Friedman while he was a member of the Statistical Research Group of Columbia University. As far as the authors are aware, no results were obtained concerning the OC and ASN functions of Friedman's procedure.

Let R_1 denote the Sequential Probability Ratio Test for testing the hypothesis that $\theta = \theta_1$ against the hypothesis that $\theta = \theta_2$. We assume for the present that either the proper constants A, B in the probability ratio test are given or that they are approximated from given α, β by the relations

$$A \sim \frac{1-\beta}{\alpha} \qquad B \sim \frac{\beta}{1-\alpha}.$$

Here α and β are upper bounds on the probabilities of first and second types of errors, respectively.

Let R_2 represent the S.P.R.T. for testing the hypothesis that $\theta = \theta_3$ against the alternative that $\theta = \theta_4$. For this test we assume that (α, β, A, B) are replaced by $(\hat{\alpha}, \hat{\beta}, \hat{A}, \hat{B})$ and as above that either $\hat{A}$ and $\hat{B}$ are given or that they are approximated from given $\hat{\alpha}$, $\hat{\beta}$.

The decision procedure is carried out as follows:

Both R_1 and R_2 are computed at each stage of the inspection until

Either: One ratio leads to a decision to stop before the other. Then the former is no longer computed and the latter is continued until it leads to a decision to stop.

Or: Both R_1 and R_2 lead to a decision to stop at the same stage. In this event both computations are discontinued.

The following table gives the rule R for the decisions to be made corresponding to all possible outcomes of R_1 and R_2.

	R_1		R_2		R
If	accepts θ_1	and	accepts θ_3	then	accepts H_1
If	accepts θ_2	and	accepts θ_3	then	accepts H_2
If	accepts θ_2	and	accepts θ_4	then	accepts H_3

We shall show that acceptance of both θ_1 and θ_4 is impossible when $(\hat{A}, \hat{B}) = (A, B)$. For this purpose we need the acceptance number and rejection number formulas. (See page 119 of [1]).

Acceptance Number *Rejection Number*

$$R_1: \frac{\sigma^2}{\Delta} \log B + a_1 n < \sum_{\alpha=1}^{n} x_\alpha < \frac{\sigma^2}{\Delta} \log A + a_1 n$$

$$R_2: \frac{\sigma^2}{\Delta} \log B + a_2 n < \sum_{\alpha=1}^{n} x_\alpha < \frac{\sigma^2}{\Delta} \log A + a_2 n.$$

We shall assume *for convenience* that "between observations" R_1 is tested before R_2 and let the term "initial decision" refer to the first decision made.

Assume θ_1 and θ_4 are both accepted. Then if θ_1 is accepted initially at the mth stage

$$\sum_{\alpha=1}^{m} x_\alpha \leq \frac{\sigma^2}{\Delta} \log B + a_1 m.$$

Since

$$\frac{\sigma^2}{\Delta} \log B + a_1 m < \frac{\sigma^2}{\Delta} \log B + a_2 m$$

it follows that θ_4 is rejected at the same stage, contradicting the hypothesis. Similarly if θ_4 is accepted initially at the mth stage, then

$$\sum_{\alpha=1}^{m} x_\alpha \geqq \frac{\sigma^2}{\Delta} \log A + a_2 m.$$

Since

$$\frac{\sigma^2}{\Delta} \log A + a_2 m > \frac{\sigma^2}{\Delta} \log A + a_1 m$$

it follows that θ_1 is rejected at the same or at an earlier stage, contradicting the assumption that the acceptance of θ_4 is an initial decision. Hence θ_1 and θ_4 cannot both be accepted.

A geometrical representation of the rule R is given in Figure 1.

R can now be described as follows: Continue taking observations until an acceptance region (shaded area) is reached or both dashed lines are crossed. In the former case, stop and accept as shown above. In the latter case stop and accept H_2.

The proof above that θ_1 and θ_4 cannot both be accepted consists of noting that a point below the acceptance line for θ_1 is already below the rejection line for θ_4 and that a point above the acceptance line for θ_4 is already above the rejection line for θ_1.

If $(\hat{A}, \hat{B}) \neq (A, B)$, a necessary and sufficient condition for the impossibility of accepting θ_1 and θ_4 is that at $n = 1$ the following inequalities should hold.

Rejection Number (of θ_1) for $R_1 \leqq$ Rejection Number (of θ_3) for R_2

and

Acceptance Number (of θ_1) for $R_1 \leqq$ Acceptance Number (of θ_3) for R_2.

In symbols

$$\frac{\sigma^2}{\Delta} \log A + a_1 \leqq \frac{\sigma^2}{\Delta} \log \hat{A} + a_2$$

and

$$\frac{\sigma^2}{\Delta} \log B + a_1 \leqq \frac{\sigma^2}{\Delta} \log \hat{B} + a_2.$$

These can be written as

$$\frac{A}{\hat{A}} \leqq e^{d\Delta/\sigma^2} \quad \text{and} \quad \frac{B}{\hat{B}} \leqq e^{d\Delta/\sigma}$$

respectively, where $d = a_2 - a_1$.

Since $\frac{d\Delta}{\sigma^2} > 0$, the above inequalities are certainly fulfilled when

$$\frac{B}{\bar{B}} \leqq 1 \quad \text{and} \quad \frac{A}{\bar{A}} \leqq 1. \tag{2.1}$$

In what follow in this paper, we shall restrict ourselves to cases where acceptance of both θ_1 and θ_4 is impossible, even if this is not stated explicitly.

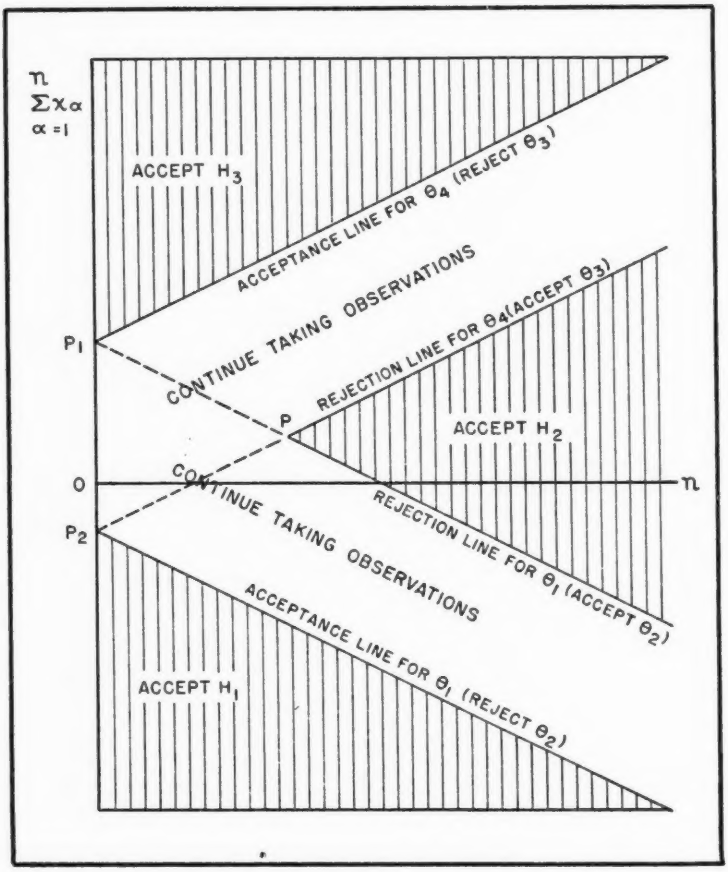

FIGURE 1

3. Derivation of OC Functions. Let $L(H_i \mid \theta, R)$ denote the probability of accepting H_i when θ is the true mean and R is the sequential rule used. Let H_{θ_i} denote the hypothesis that $\theta = \theta_i$. Since, as shown above, H_1 is accepted if and only if θ_1 is accepted, we have

$$L(H_1 \mid \theta, R) = L(H\theta_1 \mid \theta, R_1). \tag{3.1}$$

Similarly,

$$L(H_3 \mid \theta, R) = L(H_{\theta_4} \mid \theta, R_2). \tag{3.2}$$

From the fact that R_1 and R_2 each terminate at some finite stage with probability one, it follows that R will terminate at some finite stage with probability one. Hence

$$L(H_2 \mid \theta, R) = 1 - L(H_1 \mid \theta, R) - L(H_3 \mid \theta, R). \tag{3.3}$$

From pp. 50–52 of [1], the following equations are obtained.

$$L(H_1 \mid \theta, R) = L(H_{\theta_1} \mid \theta, R_1) = \frac{A^{h_1} - 1}{A^{h_1} - B^{h_1}} \tag{3.4}$$

where

$$h_1 = h_1(\theta) = \frac{\theta_2 + \theta_1 - 2\theta}{\theta_2 - \theta_1} = \frac{a_1 - \theta}{\frac{\Delta}{2}}$$

and

$$L(H_{\theta_3} \mid \theta, R_2) = \frac{\hat{A}^{h_2} - 1}{\hat{A}^{h_2} - \hat{B}^{h_2}} \tag{3.5}$$

where

$$h_2 = h_2(\theta) = \frac{\theta_4 + \theta_3 - 2\theta}{\theta_4 - \theta_3} = \frac{a_2 - \theta}{\frac{\Delta}{2}}.$$

These equations involve an approximation, as explained in [1].

Hence

$$L(H_3 \mid \theta, R) = L(H_{\theta_4} \mid \theta, R_2) = 1 - L(H_{\theta_3} \mid \theta, R_2) = \frac{1 - \hat{B}^{h_2}}{\hat{A}^{h_2} - \hat{B}^{h_2}} \tag{3.6}$$

and

$$L(H_2 \mid \theta, R) = 1 - \frac{A^{h_1} - 1}{A^{h_1} - B^{h_1}} - \frac{1 - \hat{B}^{h_2}}{\hat{A}^{h_2} - \hat{B}^{h_2}} = \frac{1 - B^{h_1}}{A^{h_1} - B^{h_1}} - \frac{1 - \hat{B}^{h_2}}{\hat{A}^{h_2} - \hat{B}^{h_2}}. \tag{3.7}$$

Since $L(H_1 \mid \theta, R) = L(H_{\theta_1} \mid \theta, R_1)$, it follows that $L(H_1 \mid \theta, R)$ is a monotonically decreasing function of θ and that

$$L(H_1 \mid -\infty, R) = 1; \qquad L(H_1 \mid \infty, R) = 0$$

$$L(H_1 \mid \theta_1, R) = 1 - \alpha; \qquad L(H_1 \mid \theta_2, R) = \beta$$

$$L(H_1 \mid a_1, R) = \frac{\log A}{\log A + |\log B|}.$$

Similarly, since $L(H_3 \mid \theta, R) = 1 - L(H_{\theta_3} \mid \theta, R_2)$, it follows that $L(H_3 \mid \theta, R)$ is a monotonically increasing function of θ and that

$$L(H_3 \mid -\infty, R) = 0; \qquad L(H_3 \mid \infty, R) = 1$$

$$L(H_3 \mid \theta_3, R) = \hat{\alpha}; \qquad L(H_3 \mid \theta_4, R) = 1 - \hat{\beta}$$

$$L(H_3 \mid a_2, R) = \frac{|\log \hat{B}|}{\log \hat{A} + |\log \hat{B}|}.$$

Since $L(H_2 \mid \theta, R) = 1 - L(H_1 \mid \theta, R) - L(H_3 \mid \theta, R)$ it follows easily from the above results that

$$L(H_2 \mid -\infty, R) = 0; \qquad L(H_2 \mid \infty, R) = 0$$

$$L(H_2 \mid \theta, R) < \alpha \quad \text{for} \quad \theta < \theta_1; \qquad L(H_2 \mid \theta, R) < \hat{\beta} \quad \text{for} \quad \theta > \theta_4$$

$$\frac{|\log B|}{\log A + |\log B|} - \hat{\alpha} < L(H_2 \mid a_1, R) < \frac{|\log B|}{\log A + |\log B|}$$

$$\frac{\log \hat{A}}{\log \hat{A} + |\log \hat{B}|} - \beta < L(H_2 \mid a_2, R) < \frac{\log \hat{A}}{\log \hat{A} + |\log \hat{B}|}$$

$$1 - \beta - \hat{\alpha} < L(H_2 \mid \theta, R) < 1 \quad \text{for} \quad \theta_2 \leqq \theta \leqq \theta_3.$$

4. Probability of Correct Decision. Denote the probability of a correct decision by $L(\theta/R)$. It is defined as follows:

Interval	*Correct Decisions*	$L(\theta \mid R)$
$\theta \leqq \theta_1$	acceptance of H_1	$L(H_1 \mid \theta, R)$
$\theta_1 < \theta < \theta_2$	acceptance of H_1 or H_2	$L(H_1 \mid \theta, R) + L(H_2 \mid \theta, R)$
$\theta_2 \leqq \theta \leqq \theta_3$	acceptance of H_2	$L(H_2 \mid \theta, R)$
$\theta_3 < \theta < \theta_4$	acceptance of H_2 or H_3	$L(H_2 \mid \theta, R) + L(H_3 \mid \theta, R)$
$\theta_4 \leqq \theta$	acceptance of H_3	$L(H_3 \mid \theta, R)$

It should be noted that at points of discontinuity, $L(\theta, \mid R)$ is defined as the smaller of the two limiting values.

We shall now discuss some monotonicity properties of the function $L(\theta \mid R)$. From the fact that $L(H_{\theta_1} \mid \theta, R_1)$ and $L(H_{\theta_3} \mid \theta, R_2)$ are continuous with continuous first and second derivatives and are monotonically decreasing for all θ with a single point of inflection in the intervals $\theta_1 < \theta < \theta_2$ and $\theta_3 < \theta < \theta_4$ respectively, it follows that

(i) $L(\theta \mid R)$ is monotonically decreasing with negative curvature for $-\infty < \theta \leqq \theta_1$.

(ii) $L(\theta \mid R)$ is monotonically increasing with negative curvature for $\theta_4 \leqq \theta < \infty$.

Making use of (3.3) we have further

(iii) $L(\theta \mid R)$ is monotonically decreasing with negative curvature for $\theta_1 < \theta < \theta_2$.

(iv) $L(\theta \mid R)$ is monotonically increasing with negative curvature for $\theta_3 < \theta < \theta_4$.

(v) For $\theta_2 \leqq \theta \leqq \theta_3$, $\frac{d}{d\theta} L(\theta \mid R) = -\left[\frac{d}{d\theta} L(H_1 \mid \theta, R) + \frac{d}{d\theta} L(H_3 \mid \theta, R)\right]$ is decreasing, since $\frac{d}{d\theta} L(H_1 \mid \theta, R)$ and $\frac{d}{d\theta} L(H_3 \mid \theta, R)$ are increasing. In other words $L(\theta \mid R)$ has negative curvature for $\theta_2 \leqq \theta \leqq \theta_3$.

In the special case when $A = \hat{A} = \frac{1}{B} = \frac{1}{\hat{B}}$ and the origin is taken at $\frac{a_1 + a_2}{2}$ for the sake of convenience, it is easy to see that $L(\theta \mid R)$ is symmetric with respect to the origin and, because of (v), has a local maximum at $\theta = 0$.

5. Choice of the constants A, B, $\hat{A}$, $\hat{B}$ to insure prescribed Lower Bounds for $L(\theta \mid R)$. We shall deal here with the question of choosing A, B, $\hat{A}$ and $\hat{B}$ such that $L(\theta \mid R) \geqq 1 - \gamma_1$ when $\theta \leqq \theta_1$, $L(\theta \mid R) \geqq 1 - \gamma_2$ when $\theta_1 < \theta < \theta_4$, and $L(\theta \mid R) \geqq 1 - \gamma_3$ when $\theta \geqq \theta_4$. From the monotonic properties of the correct decision function it is only necessary to insure that

$$(5.1) \quad L(\theta_1 \mid R) = 1 - \gamma_1, L(\theta_2 \mid R) = L(\theta_3 \mid R) = 1 - \gamma_2 \text{ and } L(\theta_4 \mid R) = 1 - \gamma_3.$$

The following relations will be needed:

$$h_1(\theta_1) = h_2(\theta_3) = 1 = -h_1(\theta_2) = -h_2(\theta_4)$$

$$h_2(\theta_2) = \frac{\theta_3 + \theta_4 - 2\theta_2}{\Delta} = \frac{d - \frac{\Delta}{2}}{\frac{\Delta}{2}} = r \quad \text{(say)}$$

$$h_1(\theta_3) = \frac{\theta_1 + \theta_2 - 2\theta_3}{\Delta} = \frac{-d + \frac{\Delta}{2}}{\frac{\Delta}{2}} = -r$$

where $d = \theta_4 - \theta_2 = \theta_3 - \theta_1 = a_2 - a_1$.

The following four equations are obtained from (5.1):

$$(5.2) \quad 1 - L(H_1 \mid \theta_1, R) = L(H_{\theta_2} \mid \theta_1, R_1) = \frac{1 - B}{A - B} = \gamma_1$$

$$(5.3) \quad \begin{aligned} 1 - L(H_2 \mid \theta_2, R) &= L(H_1 \mid \theta_2, R) + L(H_3 \mid \theta_2, R) \\ &= \frac{B(A-1)}{A - B} + \left[\frac{1 - \hat{B}^r}{\hat{A}^r - \hat{B}^r}\right] = \gamma_2 \end{aligned}$$

$$(5.4) \quad \begin{aligned} 1 - L(H_2 \mid \theta_3, R) &= L(H_3 \mid \theta_3, R) + L(H_1 \mid \theta_3, R) \\ &= \frac{1 - \hat{B}}{\hat{A} - \hat{B}} + \left[\frac{B^r(A^r - 1)}{A^r - B^r}\right] = \gamma_2 \end{aligned}$$

$$(5.5) \quad 1 - L(H_3 \mid \theta_4, R) = L(H_{\theta_3} \mid \theta_4, R_2) = \frac{\hat{B}(\hat{A} - 1)}{\hat{A} - \hat{B}} = \gamma_3.$$

The "bracketed terms" represent quantities less than $\hat{\alpha}$ and β respectively and if r is sufficiently large they can be neglected. This will be made more precise but first let us note the results of neglecting the bracketed terms.

From (5.2) and (5.3) we obtain

$$(5.6) \quad B(1 - \gamma_1) = \gamma_2, \quad \text{whence} \quad B = \frac{\gamma_2}{1 - \gamma_1}.$$

From (5.2) and (5.6)

$$A = \frac{1 - B(1 - \gamma_1)}{\gamma_1} \quad \text{whence} \quad A = \frac{1 - \gamma_2}{\gamma_1}. \tag{5.7}$$

Since the last two equations are obtained from the first two by the permutation $A \to \hat{A}$, $B \to \hat{B}$, $\gamma_1 \to \gamma_2$, $\gamma_2 \to \gamma_3$, we have

$$\hat{B} = \frac{\gamma_3}{1 - \gamma_2}$$

$$\hat{A} = \frac{1 - \gamma_3}{\gamma_2}.$$

If $\gamma_1 = \gamma_2 = \gamma_3 = \gamma$ (say) then $A = \hat{A} = \dfrac{1}{B} = \dfrac{1}{\hat{B}} = \dfrac{1 - \gamma}{\gamma}$.

We shall consider the bracketed quantities negligible if the result of neglecting them produces a change of less than 20% in $[1 - L(\theta \mid R)]$ at $\theta = \theta_2, \theta_3$ respectively, i.e., if

$$\frac{1 - \hat{B}^r}{\hat{A}^r - {}^r\hat{B}} = \frac{1 - \left(\dfrac{\gamma_3}{1 - \gamma_2}\right)^r}{\left(\dfrac{1 - \gamma_3}{\gamma_2}\right)^r - \left(\dfrac{\gamma_3}{1 - \gamma_2}\right)^r} \leqq \frac{\gamma_2}{5} \tag{5.8}$$

and

$$\frac{B^r(A^r - 1)}{A^r - B^r} = \frac{\left(\dfrac{\gamma_2}{1 - \gamma_1}\right)^r \left[\left(\dfrac{1 - \gamma_2}{\gamma_1}\right)^r - 1\right]}{\left(\dfrac{1 - \gamma_2}{\gamma_1}\right)^r - \left(\dfrac{\gamma_2}{1 - \gamma_1}\right)^r} \leqq \frac{\gamma_2}{5}. \tag{5.9}$$

Inequality, (5.9) can be written as

$$\frac{\gamma_2^r[(1 - \gamma_2)^r - \gamma_1^r]}{(1 - \gamma_2)^r(1 - \gamma_1)^r - \gamma_1^r\gamma_2^r} \leqq \frac{\gamma_2}{5}$$

or

$$(1 - \gamma_2)^r \left[\gamma_2^r - \frac{\gamma_2}{5}(1 - \gamma_1)^r\right] \leqq (\gamma_1\gamma_2)^r\left(1 - \frac{\gamma_2}{5}\right).$$

This will certainly hold if

$$\gamma_2^r \leqq \frac{\gamma_2}{5}(1 - \gamma_1)^r$$

or if

$$\left(\frac{\gamma_2}{1 - \gamma_1}\right)^r \leqq \frac{\gamma_2}{5}.$$

Assume that γ_1, γ_2 and γ_3 are each less than $\frac{1}{2}$. Then the last inequality can be

written as

$$r \geqq \frac{\log\left(\frac{5}{\gamma_2}\right)}{\log\left(\frac{1-\gamma_1}{\gamma_2}\right)}. \tag{5.10}$$

Starting with (5.8) the same relation is obtained except that γ_1 is replaced by γ_3, namely

$$r \geqq \frac{\log \frac{5}{\gamma_2}}{\log \frac{1-\gamma_3}{\gamma_2}}. \tag{5.11}$$

Let

$$k = \frac{\log \frac{5}{\gamma_2}}{\log \frac{1-\bar{\gamma}}{\gamma_2}}$$

where $\bar{\gamma}$ is the larger of γ_1 and γ_3. Then k is the larger of the right hand members of (5.10) and (5.11). Then for (5.8) and (5.9) to hold it is sufficient that

$$r \geqq k.$$

If $\gamma_2 = .05$ and $0 < \gamma_1, \gamma_3 < .1$ then k is approximately $\frac{2}{1.3} = 1.54$. If $\gamma_2 = .01$ and $0 < \gamma_1, \gamma_3 < .1$ then k is approximately $\frac{2.7}{2} = 1.35$.

We shall now investigate under what conditions the approximate solution obtained above for A, B, $\hat{A}$, $\hat{B}$ are such that acceptance of both θ_1 and θ_4 is impossible.

It follows from (2.1) that the following pair of inequalities are sufficient for the impossibility of accepting both θ_1 and θ_4:

$$\frac{A}{\hat{A}} = \frac{\gamma_2}{\gamma_1}\frac{1-\gamma_2}{1-\gamma_3} \leqq 1; \qquad \frac{B}{\hat{B}} = \frac{\gamma_2}{\gamma_3}\frac{1-\gamma_2}{1-\gamma_1} \leqq 1. \tag{5.12}$$

If $\gamma_1 \neq \gamma_3$ let the smaller and larger of the pair (γ_1, γ_3) be denoted by $\underline{\gamma}$ and $\bar{\gamma}$ respectively. Since $1 - \underline{\gamma} > 1 - \bar{\gamma}$, then

$$\frac{\gamma_2(1-\gamma_2)}{\bar{\gamma}(1-\underline{\gamma})} < \frac{\gamma_2(1-\gamma_2)}{\underline{\gamma}(1-\bar{\gamma})}$$

and we need only consider one of the two inequalities in (5.12). The condition $\gamma_2 < \underline{\gamma}$ will in general satisfy (5.12). More precisely if all the γ's are restricted to the interval (0, .1) then

$$\frac{9}{10} \leqq \frac{1-\gamma_2}{1-\underline{\gamma}} < \frac{1-\gamma_2}{1-\bar{\gamma}} \leqq \frac{10}{9}$$

and it is sufficient for the validity of (5.12) that $\gamma_2 \leqq (.9)\,\underline{\gamma}$.

If $\gamma_1 = \gamma_3 = \gamma$ (say) then the two inequalities reduce to one

$$\gamma_2^2 - \gamma_2 + \gamma - \gamma^2 \geqq 0$$

which can be written as

$$(\gamma_2 - \gamma)(\gamma_2 - 1 + \gamma) \geqq 0.$$

Since the inequality $\gamma_2 \geqq 1 - \gamma$ is impossible when all γ's are $<\frac{1}{2}$, we see that $\gamma_2 \leq \gamma$ is sufficient for the validity of (5.12) when $\gamma_1 = \gamma_3 = \gamma < \frac{1}{2}$.

There remains the problem of finding an approximate solution for equations (5.2) to (5.5) when $r < k$. Since

$$r = \frac{d - \frac{\Delta}{2}}{\frac{\Delta}{2}} = \frac{\theta_3 - \theta_2 + \frac{\Delta}{2}}{\frac{\Delta}{2}} \geqq 1$$

we merely have to consider the interval $1 \leqq r < k$.

The following approximations are used

$$\begin{aligned} &\frac{1-B}{A-B} \sim \frac{1}{A}; \quad \frac{B(A-1)}{A-B} \sim B; \quad \frac{1-\hat{B}^r}{\hat{A}^r - \hat{B}^r} \sim \frac{1}{\hat{A}^r} \\ &\frac{1-\hat{B}}{\hat{A}-\hat{B}} \sim \frac{1}{\hat{A}}; \quad \frac{B^r(A^r-1)}{A^r-B^r} \sim B^r; \quad \frac{\hat{B}(\hat{A}-1)}{\hat{A}-\hat{B}} \sim \hat{B}, \end{aligned} \tag{5.13}$$

which upon substitution yield

$$A = \frac{1}{\gamma_1} \tag{5.14}$$

$$\hat{B} = \gamma_3 \tag{5.15}$$

$$B + \frac{1}{\hat{A}^r} = \gamma_2 \tag{5.16}$$

$$\frac{1}{\hat{A}} + B^r = \gamma_2 . \tag{5.17}$$

Subtraction of (5.17) from (5.16) shows that $B = \frac{1}{\hat{A}}$ is a solution. Substituting this result back in (5.16) leads to the equation

$$B + B^r = \gamma_2 . \tag{5.18}$$

It can easily be verified that between zero and unity this equation has exactly one root. Since $1 \leqq r < \infty$, the root of the above equation lies between $\frac{\gamma_2}{2}$ and γ_2.

Taking γ_2 as a first approximation for B and substituting $\gamma_2 + \epsilon$ for B in (5.18), we obtain

$$\epsilon + (\gamma_2 + \epsilon)^r = 0.$$

Expanding $(\gamma_2 + \epsilon)^r$ in a power series in ϵ and neglecting second and higher order terms, the above equation gives

$$\epsilon \sim \frac{\gamma_2^r}{1 + r\gamma_2^{r-1}}.$$

Thus,

$$B = \frac{1}{A} \sim \gamma_2 - \frac{\gamma_2^r}{1 + r\gamma_2^{r-1}} = \frac{\gamma_2[1 + (r-1)\gamma_2^{r-1}]}{1 + r\gamma_2^{r-1}}. \tag{5.19}$$

It is necessary to investigate under what conditions the above approximate solution satisfies (5.2) to (5.5) to within a 20% error in $[1 - L(\theta/R)]$, i.e., such that

$$-\frac{\gamma_1}{5} < \frac{\gamma_1(1-B)}{1-\gamma_1 B} - \gamma_1 < \frac{\gamma_1}{5} \tag{5.20}$$

$$-\frac{\gamma_3}{5} < \frac{\gamma_3(1-B)}{1-\gamma_3 B} - \gamma_3 < \frac{\gamma_3}{5} \tag{5.21}$$

$$-\frac{\gamma_2}{5} < \frac{B(1-\gamma_1)}{1-\gamma_1 B} + \frac{B^r(1-\gamma_3^r)}{1-(\gamma_3 B)^r} - \gamma_2 < \frac{\gamma_2}{5} \tag{5.22}$$

$$-\frac{\gamma_2}{5} < \frac{B(1-\gamma_3)}{1-\gamma_3 B} + \frac{B^r(1-\gamma_1^r)}{1-(\gamma_1 B)^r} - \gamma_2 < \frac{\gamma_2}{5} \tag{5.23}$$

where for B the value in (5.19) is understood.

It can be shown that if γ_1, γ_2, γ_3, are each between zero and .1 then the inequalities (5.20) to (5.23) hold. Furthermore it can be shown that if, in addition $\gamma_2 \leqq \min(\gamma_1, \gamma_3)$ then also the inequalities (2.1) hold. The latter inequalities are sufficient to ensure the impossibility of accepting both θ_1 and θ_4.

6. Bounds for the ASN Function. First we shall derive lower bounds for the ASN function. Let $E(n/\theta, R)$ denote the expected value of n when θ is the true mean and R is the sequential rule employed. For $\theta < \theta_2$ the probability of coming to a decision first with R_2 is large and therefore

$$E(n/\theta, R) \sim E(n/\theta, R_1)\ \theta < \theta_2.$$

From the definition of R it follows that

$$E(n/\theta, R) > E(n/\theta, R_1) \qquad \text{for all } \theta.$$

Hence $E(n/\theta, R_1)$ serves as a close lower bound when $\theta < \theta_2$.

Similarly

$$E(n/\theta, R) \sim E(n/\theta, R_2) \qquad \text{for } \theta > \theta_3$$

$$E(n/\theta, R) > E(n/\theta, R_2) \qquad \text{for all } \theta.$$

Hence $E(n/\theta, R_2)$ serves as a close lower bound for $\theta > \theta_3$.

Combining the above we have

$$E(n/\theta, R) > \text{Max}\,[E(n/\theta, R_1)\ ,\ E(n/\theta, R_2)] \tag{6.1}$$

where, neglecting the excess over the boundary,

$$E(n/\theta, R_1) = \frac{L(H_{\theta_1}/\theta, R_1) \log B + L(H_{\theta_2}/\theta, R_1) \log A}{\frac{\Delta}{\sigma^2}(\theta - a_1)} \tag{6.2}$$

$$E(n/\theta, R_2) = \frac{L(H_{\theta_3}/\theta, R_2) \log \hat{B} + L(H_{\theta_4}/\theta, R_2) \log \hat{A}}{\frac{\Delta}{\sigma_2}(\theta - a_2)} \tag{6.3}$$

Formula (6.1) gives a valid lower bound over the whole range of θ, but this lower bound will not be very close in the interval (θ_2, θ_3), particularly in the neighbourhood of the mid-point $\frac{\theta_2 + \theta_3}{2}$. The authors were not able to find any simple method for obtaining a closer lower bound in this interval. The upper bound given later in this section will, however, be fairly close also in the interval (θ_2, θ_3) and can be used as an approximation to the exact value.

We shall now derive upper bounds for the ASN function. Let R_1^* be the following rule: "Continue to take observations until R_1 accepts θ_1." Since this implies the rejection of θ_4 at the same or at a previous stage, it follows that R must terminate not later than R_1^*. Hence

$$E(n/\theta, R_1^*) \geqq E(n/\theta, R). \tag{6.4}$$

As a matter of fact one can easily verify that $E(n/\theta, R_1^*) > E(n/\theta, R)$. Thus $E(n/\theta, R_1^*)$ is an upper bound for $E(n/\theta, R)$. This upper bound will be close when the probability of accepting θ_1 is high, i.e., for $\theta \leqq \theta_1$.

By the general formula

$$E(n) = \frac{E\left(\sum_{i=1}^{n} z_i\right)}{E(z)}$$

(see p. 53 [1]) we obtain, upon neglecting the excess over the boundary,

$$E(n/\theta, R_1^*) = \frac{\log B}{\frac{\Delta}{\sigma^2}(\theta - a_1)}. \tag{6.5}$$

This coincides with (6.2) when $L(H_{\theta_2}/\theta, R_1) = 0$.

Similarly, if R_2^* denotes the rule of continuing until R_2 accepts θ_4, then

$$E(n/\theta, R_2^*) > E(n/\theta, R) \tag{6.6}$$

$$E(n/\theta, R_2^*) = \frac{\log \hat{A}}{\frac{\Delta}{\sigma^2}(\theta - a_2)} \tag{6.7}$$

and this will be a close upper bound for $\theta \geqq \theta_4$.

If $A = \hat{A} = \frac{1}{B} = \frac{1}{\hat{B}}$ and if $a_1 + a_2 = 0$ the above results reduce to

$$E(n/\theta, R) \gtrsim E(n/\theta, R_1^*) = \frac{-h}{\lambda + \theta} \qquad \text{for} \qquad \theta \leqq \theta_1 \tag{6.8}$$

$$E(n/\theta, R) \gtrsim E(n/\theta, R_2^*) = \frac{h}{\theta - \lambda} \qquad \text{for} \qquad \theta \geqq \theta_4 \tag{6.9}$$

where the symbol $\gtrsim$ stands for a close inequality, and where

$$h = \frac{\sigma^2}{\Delta} \log A \qquad \text{and} \qquad \lambda = a_2 = -a_1 .$$

To establish an upper bound for the ASN function in the interval $\theta_2 < \theta < \theta_3$ we shall restrict ourself to the case where $A = \hat{A} = \frac{1}{B} = \frac{1}{\hat{B}}$. These relations are fulfilled by the approximate values of A, B, $\hat{A}$, $\hat{B}$ suggested in section 5 when $\gamma_1 = \gamma_2 = \gamma_3$ and $r \geqq k$. We shall choose the origin to be at $\frac{a_1 + a_2}{2}$, i.e., we put $\frac{a_1 + a_2}{2} = 0$. Then the vertex P of the triangle (P_1, P_2, P) in diagram 1 lies on the abscissa axis and $OP_1 = OP_2 = h$. The abscissa of the vertex P is $\frac{h}{\lambda} = N$ (say) where $\lambda = a_2 = -a_1$. Let $y = \sum_{i=1}^{N} X_i$ represent the sum of the first N observations. Let R_{23} denote the rule: "Continue until both θ_2 and θ_3 are accepted". This is tantamount to neglecting the two outer lines in diagram 1, i.e., the acceptance lines for θ_1 and θ_4. Then clearly,

$$E(n/\theta, R_{23}) > E(n/\theta, R). \tag{6.10}$$

When θ lies between θ_2 and θ_3 this inequality will be close, since the probability of crossing either of the two outer lines is then small.

However $E(n/\theta, R_{23})$ was found difficult to compute and it was necessary to consider instead the rule R'_{23}: "Take N observations. If $y = \sum_{i=1}^{N} X_i < 0$ then continue until θ_2 is accepted. If $y > 0$ then continue until θ_3 is accepted".[3] Clearly,

$$E(n/\theta, R'_{23}) > E(n/\theta, R_{23}). \tag{6.11}$$

This inequality, however, will be close only if the probability of concluding the test before N observations, given that $\theta_2 < \theta < \theta_3$, is small.

Some investigations by the authors seem to indicate that the inequality (6.11) will be close when $\Delta < \lambda$. This inequality is likely to be fulfilled in practical problems.

We shall now proceed to determine the value of $E(n/\theta, R'_{23})$. Neglecting the excess over the boundary, we have

$$E\left(n/\theta, R'_{23}, \sum_{i=1}^{N} x_i = y\right) = \frac{h}{\lambda} + \frac{y}{\lambda - \theta} \qquad \text{for} \qquad y > 0 \tag{6.12}$$

[3] The event $y = 0$ has probability zero and it is indifferent what rule is adopted for that case.

and

$$(6.13) \qquad E\left(n/\theta, R'_{23}, \sum_{i=1}^{N} x_i = y\right) = \frac{h}{\lambda} - \frac{y}{\lambda + \theta} \qquad \text{for} \qquad y < 0$$

where, for any condition C, $E(n/\theta, R, C)$ denotes the conditional expected value of n given that the true mean is θ, that R is the sequential rule used and that the condition C is fulfilled.

Multiplying with the density of y and then integrating with respect to y, we obtain after simplification

$$(6.14) \quad E(n/\theta, R'_{23}) = \frac{1}{\lambda^2 - \theta^2}\left[h\lambda + 2h\theta\phi\left(\frac{\theta}{\sigma}\sqrt{\frac{h}{\lambda}}\right) + 2\sigma\sqrt{\frac{h\lambda}{2\pi}}\, e^{-(h\theta^2/2\lambda\sigma^2)}\right]$$

where $\phi(x) = \int_0^x \frac{e^{-(y^2/2)}}{\sqrt{2\pi}}\, dy$, and $\theta_2 < \theta < \theta_3$.

In particular, for $\theta = 0$ we get

$$(6.15) \qquad E(n/\theta = 0, R'_{23}) = \frac{h}{\lambda} + \frac{\sigma}{\lambda^2}\sqrt{\frac{2h\lambda}{\pi}}.$$

To establish a close upper bound for $\theta_3 < \theta < \theta_4$ we must bring the line of acceptance of θ_4 into account. The line of acceptance of θ_1 can be disregarded since the probability of accepting θ_1 is very small.

We therefore define the rule R_{34} as follows:

"Continue with R_1 until θ_2 is accepted and with R_2 until either θ_3 or θ_4 is accepted."

Since the ASN function for R_{34} is difficult to compute we define a modified rule R'_{34} as follows:

"Proceed to take $N\left(= \frac{h}{\lambda}\right)$ observations without regard to any rule. If $y = \sum_{i=1}^{N} X_i < 0$ then continue only with R_1 until θ_2 is accepted. If $0 < y < 2h$ then continue only with R_2 until either θ_3 or θ_4 is accepted. If $y \geqq 2h$ then stop taking observations and accept H_3."

It is clear that the following inequalities hold

$$(6.16) \qquad E(n/\theta, R'_{34}) \gtrdot E(n/\theta, R_{34}) > E(n/\theta, R).$$

The proximity of $E(n/\theta, R_{34})$ and $E(n/\theta, R)$, as stated above, is based on the fact that the probability of accepting θ_1, when $\theta_3 < \theta < \theta_4$, is small.

The proximity of $E(n/\theta, R_{34})$ and $E(n/\theta, R'_{34})$ is assured if the probability of terminating with R_{34} (and with R) before N observations is small. It can be shown that the latter condition is fulfilled when $\Delta < \lambda$. In terms of the quantity r defined in Section 5 this can be written as $r > 3$.

To determine the value of $E(n/\theta, R'_{34})$ the following two preliminary results will be needed:

If $0 < y < 2h$,

$$(6.17)\qquad E\left(n/\theta, R'_{34}, \sum_{i=1}^{N} x_i = y\right) = \frac{h}{\lambda} + \frac{2h - y - 2h\left[\dfrac{1 - e^{-(2/\sigma^2)(\lambda-\theta)(2h-y)}}{1 - e^{-(4h/\sigma^2)(\lambda-\theta)}}\right]}{\theta - \lambda}$$

$$= C \text{ (say)}.$$

If $y < 0$,

$$(6.18)\qquad E\left(n/\theta, R'_{34}, \sum_{i=1}^{N} x_i = y\right) = \frac{h}{\lambda} - \frac{y}{\lambda + \theta} = D \text{ (say)}.$$

Both are easily obtained from formula (7.25) on p. 123 of [1].
Multiplying with the density of y and integrating with respect to y, we obtain after simplification

$$(6.19)\qquad \begin{aligned} E(n/\theta, R'_{34}) = \frac{h}{\lambda} &+ \left[\phi\left(\frac{2\lambda - \theta}{\sigma}\sqrt{\frac{h}{\lambda}}\right) + \phi\left(\frac{\theta}{\sigma}\sqrt{\frac{h}{\lambda}}\right)\right] \\ &\cdot \frac{h}{(\lambda - \theta)}\left(\frac{\theta}{\lambda} - \frac{2e^{-(2h(\lambda-\theta)/\sigma^2)}}{1 + e^{-2h(\lambda-\theta)/\sigma^2}}\right) \\ &+ \frac{\sigma}{\lambda(\lambda - \theta)}\sqrt{\frac{h\lambda}{2\pi}}\left[e^{-(h\theta^2/2\lambda\sigma^2)} - e^{-h(2\lambda-\theta)^2/2\lambda\sigma^2}\right] \\ &- \frac{h\theta}{2\lambda(\lambda + \theta)}\left[1 - 2\phi\left(\frac{\theta}{\sigma}\sqrt{\frac{h}{\lambda}}\right)\right] + \frac{\sigma}{\lambda(\lambda + \theta)}\sqrt{\frac{h\lambda}{2\pi}}\, e^{-(h\theta^2/2\lambda\sigma^2)}. \end{aligned}$$

Formula (6.19) is an improvement on (6.14) as it will give for any θ a smaller upper bound, but in the neighborhood of the origin the difference is insignificant.
For $\theta = \lambda$ we obtain from (6.19) using L'Hopital's rule

$$(6.20)\qquad \begin{aligned} E(n/\lambda, R'_{34}) = \frac{h^2}{\sigma^2} &- \frac{h}{4\lambda\sigma^2}(4h\lambda - 3\sigma^2) \\ &\cdot\left[1 - 2\phi\left(\frac{\sqrt{h\lambda}}{\sigma}\right)\right] + \left(\frac{h\lambda + \sigma^2}{2\lambda^2\sigma}\right)\sqrt{\frac{h\lambda}{2\pi}}\, e^{-(h\lambda/2\sigma^2)}. \end{aligned}$$

If $\dfrac{\sqrt{h\lambda}}{\sigma} > 2.5$, the above formula can be approximated by

$$(6.21)\qquad E(n/\lambda, R'_{34}) \sim \frac{h^2}{\sigma^2} + \frac{2\sigma}{\lambda^2}\sqrt{\frac{h\lambda}{2\pi}}\, e^{-(h\lambda/2\sigma^2)}.$$

Since the right hand member above lies between $\dfrac{h^2}{\sigma^2}$ and (1.002) $\dfrac{h^2}{\sigma^2}$ when $\dfrac{\sqrt{h\lambda}}{\sigma} > 2.5$ then for practical purposes

$$(6.22)\qquad E(n/\lambda, R'_{34}) \sim \frac{h^2}{\sigma^2} \qquad \text{when } \left(\frac{\sqrt{h\lambda}}{\sigma} > 2.5\right).$$

An upper bound for $E(n/\theta, R)$ for $\theta_1 < \theta < \theta_2$ can be obtained by defining R_{12} and R'_{12} in an analagous way to R_{34} and R'_{34}. Because of reasons of symmetry, $E(n/\theta, R'_{12})$ can be obtained from (6.19) by replacing θ by $-\theta$.

The method used for obtaining upper bounds for $E(n/\theta, R)$ can easily be extended to the more general case when the equalities $A = \hat{A} = \frac{1}{B} = \frac{1}{\hat{B}}$ do not necessarily hold. However, the resulting formulas are more cumbersome and we shall merely give without proof the upper bound corresponding to (6.14). This upper bound becomes

$$E(n/\theta, R'_{23}) = N + \left(\frac{N\theta - h_3}{\lambda - \theta}\right)\left[\frac{1}{2} - \phi(a)\right] + \left(\frac{h_3 - N\theta}{\lambda + \theta}\right)\left[\frac{1}{2} - \phi(b)\right] + \sigma\sqrt{\frac{N}{2\pi}}\left[\frac{e^{-a^2/2}}{\lambda - \theta} + \frac{e^{-b^2/2}}{\lambda + \theta}\right]$$

where

$$h_{11} = \frac{\sigma^2}{\Delta}\log A \qquad h_{10} = \frac{\sigma^2}{\Delta}\log B$$

$$h_{21} = \frac{\sigma^2}{\Delta}\log \hat{A} \qquad h_{20} = \frac{\sigma^2}{\Delta}\log \hat{B}$$

$$a_2 = -a_1 = \lambda$$

$$N = \frac{h_{11} - h_{20}}{2\lambda}; \qquad a = \frac{h_3 - N\theta}{\sigma\sqrt{N}}; \qquad b = \frac{h_3 + N\theta}{\sigma\sqrt{N}}; \qquad h_3 = \frac{h_{11} + h_{20}}{2}.$$

7. An Example. We shall consider the following example

$\sigma^2 = 1$, $\theta_1 = -\frac{5}{16}$, $\theta_2 = -\frac{3}{16}$, $\theta_3 = \frac{3}{16}$, $\theta_4 = \frac{5}{16}$, $\gamma_1 = \gamma_2 = \gamma_3 = \gamma = .029$

then

$$A = \hat{A} = \frac{1}{B} = \frac{1}{\hat{B}} = \frac{1-\gamma}{\gamma} = 33.5 \qquad r = 7 \gg 3 > k \sim 1.47$$

and

$$h = \frac{\sigma^2}{\Delta}\log A = 28, \lambda = \frac{\theta_3 + \theta_4}{2} = \tfrac{1}{4}, \Delta = \theta_2 - \theta_1 = \theta_4 - \theta_3 = \tfrac{1}{8}.$$

Using formulas (6.1) and (6.7) the following upper and lower bounds were obtained

θ	$\frac{5}{16}$	$\frac{6}{16}$	$\frac{7}{16}$	$\frac{8}{16}$	$\frac{9}{16}$	$\frac{10}{16}$	$\frac{12}{16}$	$\frac{14}{16}$	$\frac{16}{16}$	$\frac{18}{16}$	$\frac{20}{16}$
Upper bound...........	448	224	149	112	89.6	74.7	56	44.8	37.3	32	28
Lower bound...........	421	224	149	112	89.6	74.7	56	44.8	37.3	32	28

Formulas (6.14) and (6.1) yield

θ	0	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$
Upper Bound	146	163	229	450
Lower Bound	112	149	224	421

In the neighborhood of the origin the true value is very nearly the upper bound. From formulas (6.19), (6.22) and (6.1) we obtain

θ	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{5}{16}$
Upper Bound	422	784.5	423
Lower Bound	421	784	421

As shown above for the end points of the indifference zone, (6.19) gives better results than (6.14) or (6.7). This is as it should be since (6.19) takes into account possibilities omitted in (6.14) and (6.7). The greater accuracy of (6.19) is offset by a slight increase in computation.

In the graph of the Bounds of the ASN function shown in Figure 2, a single curve is shown wherever the upper and lower bound are sufficiently close to each other.

Since (6.14) contains an even function of θ and since elsewhere the corresponding bounds are mirror images with respect to $\theta = 0$, the bounds for negative θ are exactly the same as those for the corresponding positive θ.

Consider the following non-sequential rule applied to our problem. With a fixed number N_0 of observations compute the mean $\bar{x}$ and accept H_1 if $\bar{x}$ falls in the interval $(-\infty, a_1)$, accept H_2 if $\bar{x}$ falls in $[a_1, a_2]$ and accept H_3 if $\bar{x}$ falls in (a_2, ∞). This is certainly a reasonable procedure. One can also verify that no other non-sequential rule exists that is uniformly better (for all possible values of θ) than the one under consideration.

The two decision procedures become comparable if we introduce the indifference zones and define a wrong decision in the non-sequential case exactly as was done for our sequential procedure (see Section 1).

For the non-sequential case (just as in the sequential case) the probability of a wrong decision will be discontinuous at θ_1, θ_2, θ_3 and θ_4. At each of these points there will be a left-sided and right-sided limit, different from each other. As in the sequential case we shall take the probability of a wrong decision at a discontinuity point to be equal to the larger of the left and right hand limits. One can easily verify that the maximum probability of a wrong decision occurs at $\theta = \theta_3$ (which is equal to the value at $\theta = \theta_2$).

We then determine N_0 by setting the maximum probability of a wrong decision equal to γ, i.e.

$$\phi\left(\frac{d - \Delta/2}{\sigma}\sqrt{N_0}\right) + \phi\left(\frac{\Delta}{2\sigma}\sqrt{N_0}\right) = 1 - \gamma. \tag{7.1}$$

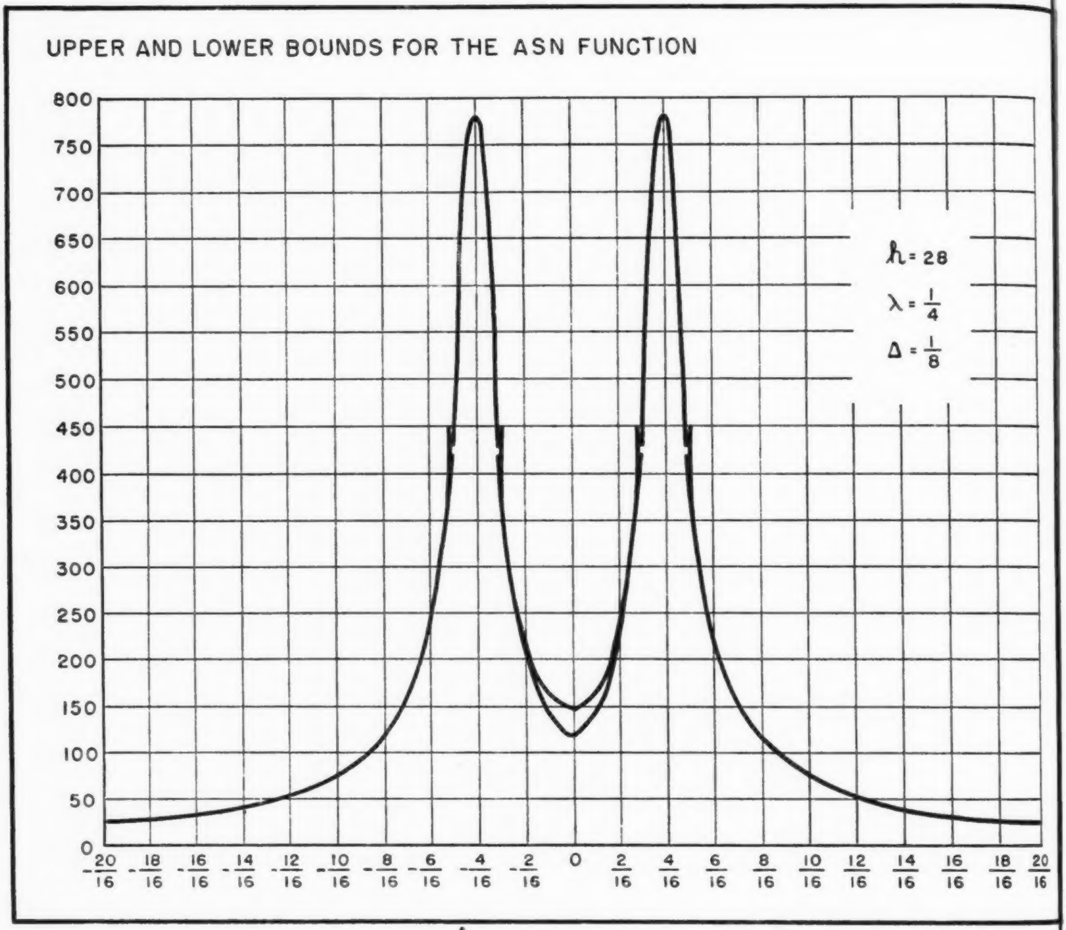

FIGURE 2

For the particular problem considered above, this gives $N_0 = 915.4$. Hence 916 observations are required in order to ensure that this non-sequential procedure will have the maximum probability $\gamma = .029$ of a wrong decision. This is to be compared with the maximum over all θ of the ASN function in the sequential procedure, which was 784.5.

Returning to (7.1) we shall derive lower and upper bounds for the root of that equation. Since

$$\infty > \frac{d - \Delta/2}{\sigma}\sqrt{N_0} \geqq \frac{\Delta}{2\sigma}\sqrt{N_0}$$

it is clear that the root of the equation

$$\phi\left(\frac{\Delta}{2\sigma}\sqrt{N_0}\right) + \phi\left(\frac{\Delta}{2\sigma}\sqrt{N_0}\right) = 1 - \gamma$$

is an upper bound for the root of (7.1) and that the root of the equation

$$\phi(\infty) + \phi\left(\frac{\Delta}{2\sigma}\sqrt{N_0}\right) = 1 - \gamma$$

or

$$\phi\left(\frac{\Delta}{2\sigma}\sqrt{N_0}\right) = \frac{1}{2} - \gamma$$

is a lower bound for the root of (7.1). We shall compare the value of $x = \frac{\Delta}{2\sigma}\sqrt{N_0}$ with the value of $y = \frac{\Delta}{2\sigma}\sqrt{\operatorname*{Max}_{\theta}\text{ ASN}}$. Since

$$\operatorname*{Max}_{\theta}\text{ (ASN function)} \sim \frac{h^2}{\sigma^2} = \frac{\sigma^2}{\Delta^2}\left(\log\frac{1-\gamma}{\gamma}\right)^2 \text{ (for sufficiently small } \frac{\Delta}{d}\text{)}.$$

then

$$y = \frac{\Delta}{2\sigma}\sqrt{\operatorname*{Max}_{\theta}\text{ ASN}} \sim \frac{1}{2}\log\frac{1-\gamma}{\gamma} \text{ (for sufficiently small } \frac{\Delta}{d}\text{)}.$$

The following table gives upper and lower bounds for x and the corresponding value of y for the type of example under consideration, i.e., when $A = \hat{A} = \frac{1}{B} = \frac{1}{\hat{B}}$ and $r \geqq k$.

γ	.001	.002	.005	.008	.01	.05	.1
$\underline{x}$ and $\bar{x}$	3.08–3.31	2.87–3.10	2.57–2.81	2.41–2.65	2.33–2.58	1.64–1.96	1.28–1.65
y	3.45	3.11	2.65	2.41	2.30	1.47	1.10

As the table shows[4] for $.1 > \gamma > .008$

$$x > \underline{x} > y$$

[4] Actually, the inequality in question is shown only for the values of γ given in the table. However it can be verified that the inequality remains valid for all values of γ between .1 and .008.

and hence

$$N_0 > \operatorname*{Max}_{\theta} \text{ASN} \qquad \left(\text{for sufficiently small } \frac{\Delta}{d}\right).$$

The statement and the table above are not meant to delimit the region in which the sequential rule is superior to the non-sequential procedure.

REFERENCES

[1] Abraham Wald, *Sequential Analysis*, John Wiley and Sons, 1947.

[2] P. Armitage, "Some sequential tests of student's hypothesis," *Supplement to the Journal of the Royal Statistical Society, Vol.* 9 (1947) No. 2, p. 250.

MOMENTS OF RANDOM GROUP SIZE DISTRIBUTIONS[1]

By John W. Tukey

Princeton University

1. Summary. A number of practical problems involve the solution of a mathematical problem of the class described in the classical language of probability theory as follows: "A number of balls are *independently* distributed among a number of boxes, how many boxes contain no balls, 1 ball, 2 balls, 3 balls, and so on." Problems arising in the oxidation of rubber and the genetics of bacteria are discussed as applications.

A method is given of solving problems of this sort when "how many" is adequately answered by the calculation of means, variances, covariances, third moments, etc. The method is applied to a number of the simplest cases, where the number of balls is fixed, binomially distributed or Poisson and where the "sizes" of the boxes are equal or unequal.

2. Introduction. The distribution of the number of empty boxes has been investigated by Romanovsky in 1934 [3], and, apparently independently, by Stevens in 1937 [4]. Romanovsky investigated the case of N equal boxes and m balls for (i) the case where the balls are independent, and (ii) the case where there is a limit to the size of each box. He gives no motivation for the problem, and shows that certain limiting distributions approach normality. Stevens investigated the case of m independent balls for N boxes (i) of equal size, and (ii) of unequal size, and developed a useful approximation for the last case. Stevens was concerned with this problem in order to test box counts for nonrandomness by comparing the number of empty boxes with expectation. The reader interested in that problem is referred to his paper.

The results derived in Part II are based on the use of a chance generating function, a technique which applies easily to the case where the balls are independent. Thus Romanovsky's results for the case of boxes of limited size are neither included or extended. For the other cases where the number of empty boxes has been considered, the results below seem to provide simple moments and cross-moments for the numbers of boxes with any number of balls to the extent previously available for the number of empty boxes. Both Romanovsky and Stevens investigated the actual distribution of the number of empty boxes. A similar investigation of the distribution of the number of b-ball boxes has *not* been carried out here.

3. A chemical problem. In studying the oxidation of rubber, Tobolsky and coworkers were led to propose the following problem: "If a mass of rubber originally consisted of N chains of equal length, if each chain can be broken at a

---

[1] Prepared in connection with research sponsored by the Office of Naval Research.

large number of places by the reaction with one oxygen molecule, if there are m oxygen molecules each equally likely to react at each link, and if mNp molecules have reacted, what is the probable number of original chains which are now in $b + 1$ parts as a result of b oxygen molecules having reacted with b of their links?

Here an original chain plays the role of a box and an oxygen molecule the role of a ball. The sort of numbers which may be taken as characteristic are:

$$\begin{aligned} N &= 10^{18} && \text{(number of chains)}, \\ m &= 10^{16} \text{ to } 10^{20} && \text{(number of oxygen molecules)}, \\ mp &= 0.01 \text{ to } 100 && \text{(average breaks/chain)}. \end{aligned}$$

Thus it is almost certainly going to be appropriate to use the results obtained by assuming N and m very large and $p = 1/N$ very small. We shall return to this example after discussing the general results.

4. A bacteriological problem. The experiments of Newcombe [1] on the irradiation and mutation of bacteria have prompted Pittendrigh to propose the following problem: "Suppose a large number of bacteria each contain m enzyme particles, which have been formed by the action of a nuclear gene. Suppose that irradiation destroys the nuclear gene in a certain fraction of the bacteria. Suppose three generations to occur, during which the m original enzyme particles are randomly distributed among the 8 descendants of an original bacterium. If a bacterium without either nuclear gene or enzyme particle is a recognizable mutant, what is the expected distribution of "families" with 0, 1, 2, 3, $\cdots$, 8 mutants?"

Here the enzyme particles are the balls, and the 8 descendants are the N boxes. We are interested in the number of empty boxes—the problem is that discussed by both Romanovsky and Stevens, with the exception of an allowance for cases where the nuclear gene was not lost. We shall return to this problem also after discussing the general results.

5. The case of large numbers. In case the number of "balls" and "boxes" is large, it is natural and has been customary in similar problems to replace discrete variables by continuous, and derive differential equations. The process runs as follows: Let $y_0, y_1, y_2, \cdots, y_b, \cdots$ be the *fractions* of the total number of boxes containing no, one, two, $\cdots$, b, $\cdots$ balls. Let t be the average number of balls per box (artificially made continuous, so that we may, for example, have a total of $13 + 3\pi$ balls). Increase t to $t + dt$, then of the y_0 boxes previously containing no balls, $y_0\,dt$ will receive one. Of the y_1 boxes previously containing one ball each, $y_1\,dt$ will receive a second, and so on. Hence

$$\frac{dy_0}{dt} = -y_0,$$

$$\frac{dy_1}{dt} = y_0 - y_1,$$

$$\cdots$$

$$\frac{dy_b}{dt} = y_{b-1} - y_b,$$

$$\cdots$$

and if we start, when $t = 0$, with $y_0 = 1$, and $y_b = 0$ for $b > 0$, we find

$$y_b = \frac{t^b}{b!} e^{-t}, \qquad b = 0, 1, 2, \cdots. \tag{1}$$

The usefulness of this result has sometimes been in doubt, thus Opatowski [2, p. 164] says in a similar connection: "Consequently ⋯ the theory appears less accurate for small values of t."

It is shown in Part II that; where n_b boxes out of the total of N contain exactly b balls: (I) When the number of *balls and boxes* is large and fixed, (1) is a good approximation to the expectation of n_b/N. (II) When the total number of balls has a Poisson distribution, and t is interpreted as the expected number, (1) *reproduces the expectation* exactly. Since it is appropriate in most problems involving chemical reactions or irradiation to take the number of balls as having a

TABLE 1

A fixed or binomial number of balls and equal boxes

HYPOTHESIS

A total of m balls are independently distributed into N boxes or elsewhere, the chance of a particular ball entering a particular box is p. The number of boxes each containing exactly b balls is n_b.

$$\text{Mean of } n_b = E(n_b) = N\binom{m}{b}(1-p)^m\left(\frac{p}{1-p}\right)^b$$

$$\text{Variance of } n_b = E(n_b)(1 - (1 - \Phi(b, b))E(n_b))$$

$$\text{Covariance of } n_b \text{ and } n_c = -(1 - \Phi(b, c))E(n_b)E(n_c)$$

$$\Phi(b, c) = \left(1 - \frac{1}{N}\right)\frac{(m-c)^{(b)}}{m^{(b)}}\left(1 - \left(\frac{p}{1-p}\right)^2\right)^m\left(\frac{1-p}{1-2p}\right)^{b+c}$$

where $m^{(b)} = m(m-1)\cdots(m-b+1)$ involving b factors

Higher moments See Section 14

$$\text{Mean of } n_0 = N(1-p)^m$$

$$\text{Mean of } n_1 = Nm(1-p)^m\left(\frac{p}{1-p}\right)$$

$$\text{Variance of } n_0 = N(1-p)^m - N^2(1-p)^{2m} + N(N-1)(1-2p)^m$$

$$\text{Variance of } n_1 = N(N-1)m(m-1)(1-2p)^{m-2}p^2 + Nm(1-p)^{m-1}p - N^2m^2(1-p)^{2m-2}p^2$$

$$\text{Covariance of } n_0 \text{ and } n_1 = N(N-1)m(1-2p)^{m-1}p - n^2m(1-p)^{2m-1}p$$

Poisson distribution, the caution suggested by (I) is often shown unnecessary by (II). For this type of problem the differential equation is entirely adequate!

It is further shown in Part II that, in the Poisson case, the second moments are exactly those which correspond to random sampling from an infinite population with the fractions indicated by the mean number of boxes with $0, 1, 2, \cdots, b, \cdots$ balls. This result is not accidental, and it is shown in Part III how we can see directly that the whole distribution in this case is that of random sampling from such a population.

6. The case of small numbers. The results of Part II also allow us to state the means, variances, and covariances, for the cases where the differential equations do not apply. The results are set forth in the following tables: Tables 1 and 2 apply to the cases where m balls are distributed among the given boxes and possibly others. Thus the total number of balls in the given boxes is either fixed, when there are no other boxes, or follows a binomial distribution.

TABLE 2

A fixed or binomial number of balls and unequal boxes

HYPOTHESIS

A total of m balls are independently distributed into N boxes or elsewhere, the chance of a particular ball entering the ith box being p_i. The average of the $p_i = p$. The sum of the squared fractional deviations of p_i from p is Λ. $p_i = p(1 + \lambda_i), \Sigma_i \lambda_i^2 = \Lambda$. Terms in $\Sigma_i\lambda_i^3$, $\Sigma_i\lambda_i^4$, etc. are to be neglected. The number of boxes each containing exactly b balls is n_b.

Mean of $n_b = E(n_b) = N\binom{m}{b}(1-p)^{m-b}p^b$ times

$$\left\{\left(1 + \frac{\Lambda}{2N(1-p)^2}\right)((mp-b)^2 - (m-b)p^2 - b(1-p)^2)\right\}$$

Variances and covariances as in Table 1, using

$$\Phi(b, c) \approx \left(1 - \frac{1}{N}\right)\frac{(m-c)^{(b)}}{m^{(b)}},\left(1 - \left(\frac{p}{1-p}\right)^2\right)^m \left(\frac{1-p}{1-2p}\right)^{b+c}\left(1 + \frac{\Lambda\psi}{2N}\right)$$

where $\psi = 2bc\left(2p - \frac{1}{N}\right)$ + terms in p^2 and in $\frac{p}{N}$

The exact value of ψ is given in Section 16.

Mean of $n_0 = N(1-p)^m\left(1 + \frac{\Lambda p^2 m(m-1)}{2N(1-p)^2}\right)$

Mean of $n_1 = Nm(1-p)^{m-1}p\left(1 + \frac{\Lambda(m-1)p(1-mp)}{2N(1-p)^2}\right)$

TABLE 3

Poisson balls and equal boxes

HYPOTHESIS

A number of balls with the Poisson distribution, and expectation Nt are independently placed in N boxes. The number of boxes each containing exactly b balls is n_b.

$$\text{Mean of } n_b = E(n_b) = N \frac{t^b}{b!} e^{-t}$$

$$\text{Variance of } n_b = N\left(\frac{t^b}{b!} e^{-t}\right)\left(1 - \frac{t^b}{b!} e^{-t}\right)$$

$$\text{Covariance of } n_b \text{ and } n_c = -N\left(\frac{t^b}{b!} e^{-t}\right)\left(\frac{t^c}{c!} e^{-t}\right)$$

$$\begin{aligned} \text{Mean of } n_0 &= Ne^{-t}, \\ \text{Mean of } n_1 &= Nte^{-t}, \\ \text{Variance of } n_0 &= Ne^{-t}(1 - e^{-t}), \\ \text{Variance of } n_1 &= Nte^{-t}(1 - te^{-t}), \\ \text{Covariance of } n_0 \text{ and } n_1 &= -Nte^{-2t}. \end{aligned}$$

7. Discussion of the chemical problem. The number of oxygen molecules which have reacted in a given time is, at best, distributed Poisson. Thus the differential equations would give the expected number of cuts, even if the number of balls or boxes were not large.

The fact that the numbers of balls and boxes, are large makes the variances and covariances so small as to be practically unimportant. Thus, for example, with $N = 10^{18}$, $t = 1$ (1 break per chain), we have:

$$\text{mean of } n_0 = \frac{1}{e} \times 10^{18},$$

$$\text{mean of } n_1 = \frac{1}{e} \times 10^{18},$$

$$\text{variance of } n_0 = \frac{1}{e}\left(1 - \frac{1}{e}\right) \times 10^{18},$$

$$\text{variance of } n_1 = \frac{1}{e}\left(1 - \frac{1}{e}\right) \times 10^{18},$$

$$\text{covariance of } n_0 \text{ and } n_1 = -\frac{1}{e^2} \times 10^{18}.$$

Thus the standard deviations are less than 1 part in 100 million of the mean.

TABLE 4

Poisson balls and varied boxes

HYPOTHESIS

A number of balls with the Poisson distribution are independently placed in N unequal boxes. The expected number placed in the ith box is t_i. The average of the t_i is t, $t_i = t(1 + \lambda_i)$ and $\Sigma_i \lambda_i^2 = \Lambda$. Terms in $\Sigma_i \lambda_i^3$, $\Sigma_i \lambda_i^4$, etc. are to be neglected. The number of boxes each containing exactly b balls is n_b.

$$\text{Mean of } n_b = E(n_b) = N\frac{t^b}{b!}e^{-t}\left(1 + \frac{\Lambda}{2N}((b-t)^2 - b)\right)$$

$$\text{Variance of } n_b = E(n_b) - \frac{1}{N}\left(1 + \frac{\Lambda}{2N}(b-t)^2)(E(n_b))^2\right)$$

$$\text{Covariance of } n_b \text{ and } n_c = -\frac{1}{N}\left(1 + \frac{\Lambda}{2N}((b-t)(c-t))E(n_b)E(n_c)\right)$$

$$\text{Mean of } n_0 = Ne^{-t}\left(1 + \frac{\Lambda t^2}{2N}\right)$$

$$\text{Mean of } n_1 = Nte^{-t}\left(1 + \frac{\Lambda(t^2 - 2t)}{2N}\right)$$

$$\text{Variance of } n_0 = Ne^{-t}\left(1 + \frac{\Lambda t^2}{2N}\right) - Ne^{-2t}\left(1 + \frac{3\Lambda t^2}{2N}\right)$$

$$\text{Variance of } n_1 = Nte^{-t}\left(1 + \frac{\lambda(t^2 - 2t)}{2N}\right) - Nt^2e^{-2t}\left(1 + \frac{\Lambda(3t^2 - 6t\ 1)}{2N}\right)$$

$$\text{Covariance of } n_0 \text{ and } n_1 = -Nt^2e^{-2t}\left(1 + \frac{\Lambda(3t^2 - 3t)}{2N}\right)$$

8. Discussion of the bacteriological example. Although this example came from an irradiation experiment, we are not entitled to jump to the Poisson case. The balls are not actions of radiation, but rather previously existing enzyme particles. The purpose of the radiation is merely to make a failure to hand down a particle obvious.

For simplicity, let us begin by assuming that the irradiation has been strong enough to knock out all the nuclear genes and none of the enzyme particles. We face the following problem: "If the m enzyme particles are divided by chance among 8 descendants, what should be the distribution of mutants, that is, of boxes with no balls?"

As far as mean and variance, we can answer this question from Table 1, with $N = 8$ and $p = \frac{1}{8}$.

The results are

$$\text{mean number of mutants} = E(n_0) = 8(\tfrac{7}{8})^m,$$

$$\text{variance of same} = 8(\tfrac{7}{8})^m - 64(\tfrac{7}{8})^{2m} + 56(\tfrac{6}{8})^m.$$

For small values of m we get the values tabled below:

TABLE 5

Blanks out of 8

m	mean	variance	$\text{mean}\left(1 - \frac{\text{mean}}{8}\right)$
0	8	0.000	0.000
1	7	0.000	0.875
2	6.125	.109	1.436
3	5.359	.262	1.769
4	4.689	.417	1.941
5	4.103	.556	1.998
6	3.590	.666	1.979
7	3.142	.747	1.908
8	2.749	.799	1.804
9	2.405	.825	1.682
10	2.105	.829	1.551
15	1.079	.663	.934
20	0.554	.426	.515

We notice that the variance is substantially less than the mean.

Now it might be that the number of enzyme particles is not constant from bacterium to bacterium. It would not be unreasonable if it had a Poisson distribution. If this were the case, we would revert to the differential equation solution, which is also given in Table 3. The last column in Table 5 shows the variance which would then arise for the same means. The variance is still somewhat less than the mean. The situation is shown graphically in Figure 1.

If the actual distribution of n_0 is desired, then it can be calculated for the case where m is fixed from the tables in Stevens' paper [4], and when m is distributed Poisson it is merely a binomial distribution.

PART II

DERIVATIONS

9. The chance generating function. We are considering the following class of problems: "balls" are placed *independently* in "boxes" and then the number n_0 of empty compartments, the number n_1 of compartments containing exactly

one ball, $\cdots$, the number n_b of boxes with exactly b balls, and so on, are observed. We are interested in the moments of $n_0, n_1, n_2, \cdots, n_b, \cdots$ both simple and mixed.

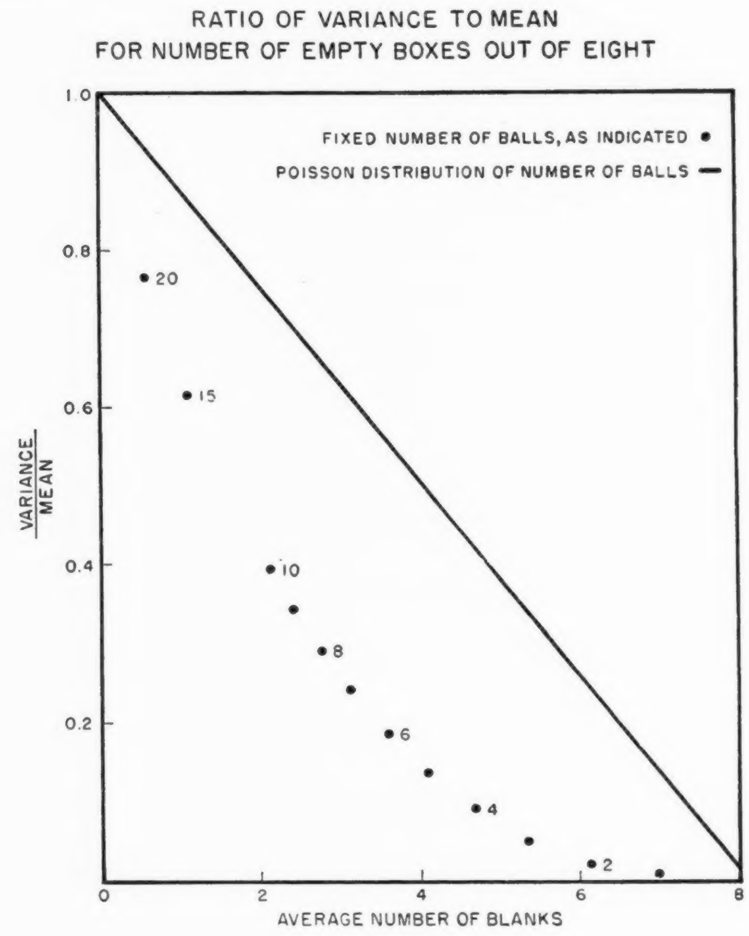

Figure 1

We define chance quantities x_{iq} by

$$x_{iq} = \begin{cases} x, & q\text{th ball in the } i\text{th box}, \\ 1, & \text{otherwise.} \end{cases}$$

Clearly the product of all x_{iq} for fixed i is given by

$$\Pi_q x_{iq} = x^{\text{(number of balls in the } i\text{th box)}}$$

Thus $\Pi_q x_{iq} = x^b$ if and only if there are exactly b balls in the ith box. Hence the coefficient of x^b in $\Sigma_i \Pi_q x_{iq}$, the sum of $\Pi_q x_{iq}$ over all boxes i, is n_b, the number of boxes containing exactly b balls.

We have the relation

$$\Sigma_b n_b x^b = f(x) = \Sigma_i \Pi_q x_{iq},$$

where $f(x)$ is a chance function, and the n_b and the x_{iq} are chance quantities.

Now we take expectations of both sides, and use the fact that the expectation of a sum is the sum of the expectations to obtain

$$\Sigma_b x^b E(n_b) = E(f(x)) = \Sigma_i E(\Pi_q x_{iq}).$$

Now x_{iq} and x_{ir}, for $q \neq r$, are independent since they are determined by different and independent balls. Hence $E(\Pi_q x_{iq}) = \Pi_q(Ex_{iq})$ and we have the basic formula

$$(1) \qquad E(f(x)) = \Sigma_b x^b E(n_b) = \Sigma_i \Pi_q E(x_{iq}).$$

10. Higher moments. By extending this device, we can obtain generating functions for higher moments. Instead of the x_{iq}, we introduce a whole sequence of chance quantities $x_{iq}, y_{iq}, z_{iq}, \cdots, w_{iq}$, defined by

$$(x_{iq}, y_{iq}, \cdots, w_{iq}) = \begin{cases} (x, y, \cdots, w), & q\text{th ball in } i\text{th box}, \\ (1, 1, \cdots, 1), & \text{otherwise}. \end{cases}$$

We find immediately that

$$\begin{aligned} f(x)f(y) \cdots f(w) &= (\Sigma_i \Pi_q x_{iq})(\Sigma_j \Pi_r y_{jr}) \cdots (\Sigma_n \Pi_p w_{np}) \\ &= \Sigma_i \Sigma_j \cdots \Sigma_n \Pi_q x_{iq} y_{jq} \cdots w_{nq}. \end{aligned}$$

Taking expectations on both sides

$$\begin{aligned} E(f(x)f(y) \cdots f(w)) &= \Sigma_i \Sigma_j \cdots \Sigma_n E(\Pi_q x_{iq} y_{jq} \cdots w_{nq}) \\ &= \Sigma_i \Sigma_j \cdots \Sigma_n \Pi_q E(x_{iq} y_{jq} \cdots w_{nq}), \end{aligned}$$

where we have used the fact that $x_{iq} y_{jq} \cdots w_{nq}$ and $x_{ir} y_{jr} \cdots w_{nr}$ are independent when $q \neq r$ since they are determined by different and independent balls.

On the other hand,

$$\begin{aligned} f(x)f(y) \cdots f(w) &= (\Sigma_b n_b x^b)(\Sigma_c n_c y^c) \cdots (\Sigma_a n_a w^a) \\ &= \Sigma_b \Sigma_c \cdots \Sigma_a (n_b n_c \cdots n_a)(x^b y^c \cdots w^a) \end{aligned}$$

so that

$$E(f(x)f(y) \cdots f(w)) = \Sigma_b \Sigma_c \cdots \Sigma_a (x^b y^c \cdots w^a) E(n_b n_c \cdots n_a).$$

Equating the two expressions for the expectation of $f(x)f(y) \cdots f(w)$, we have, finally, the generating function for $E(n_b n_c \cdots n_n)$ in the form

$$(2) \qquad \sum_{b,c,\cdots,a} (x^b y^c \cdots w^a) E(n_b n_c \cdots n_a) = \sum_{i,j,\cdots,n} \Pi_q E(x_{iq} y_{jq} \cdots w_{nq}).$$

Thus a knowledge of $E(x_{iq} y_{jq} \cdots w_{nq})$ will allow us to determine the moments of the n's.

11. A fixed or binomial number of balls and equal boxes. Let there be N boxes, and m balls, each with probability p of entering each box. If $pN = 1$ we have the case where m balls always appear in the boxes taken together—the case of a fixed number of balls. If $pN < 1$, the number of balls appearing in all boxes taken together is a binomial with expectation mpN.

Now x_{iq} equals 1 with probability $1 - p$ and equals x with probability p, hence (1) becomes

$$\Sigma_b x^b E(n_b) = \Sigma_i \Pi_q (1 - p + px) = N(1 - p + px)^m.$$

Using the binomial theorem, the coefficient of x^b is

$$E(n_b) = N\binom{m}{b}(1 - p)^{m-b}p^b = N\binom{m}{b}(1 - p)^m\left(\frac{p}{1 - p}\right)^b. \tag{3}$$

Now if p is small, we may approximate $1 - p$ by e^{-p} and by 1, respectively, in its two occurences, where

$$E(n_b) \approx N\binom{m}{b}e^{-mp}p^b$$

and if m is large compared to b this becomes

$$E(n_b) \approx N\frac{(mp)^b}{b!}e^{-mp}.$$

12. Second moments. We must study $E(x_{iq}y_{jq})$. If $i = j$ then this is $(1 - p + pxy)$ since the qth ball falls into *both* the ith and jth boxes with probability p, otherwise into neither. If $i \neq j$, we immediately find the expectation to be $(1 - 2p + px + py)$.

Hence, since $i = j$ in N cases, and $i \neq j$ in $N(N - 1)$ cases,

$$\Sigma_{ij}\Pi_q E(x_{iq}y_{jq}) = N(1 - p + pxy)^m + N(N - 1)(1 - 2p + px + py)^m,$$

by (2) this equals $\Sigma_{bc}x^b y^c E(n_b n_c)$, and using the multinomial expansion we find

$$E(n_b n_c) = N(N - 1)\binom{m}{b\,c}(1 - 2p)^{m-b-c}p^{bc} + \delta(b, c)N\binom{m}{b}(1 - p)^{m-b}p^b,$$

where $\delta(b, c) = 1$ when $b = c$ and is zero otherwise, and where the multinomial coefficient $\binom{m}{b\,c}$ is given by

$$\binom{m}{b\,c} = \frac{m!}{b!c!(m - b - c)!}.$$

We now set

$$E(n_b n_c) = E(n_b)E(n_c)\Phi(b, c) + \delta(b, c)E(n_b), \tag{4}$$

when

$$\Phi(b, c) = \frac{N(N-1)\binom{m}{b\,c}(1-2p)^m\left(\dfrac{p}{1-2p}\right)^{b+c}}{N\binom{m}{b}(1-p)^m\left(\dfrac{p}{1-p}\right)^{b} N\binom{m}{c}(1-p)^m\left(\dfrac{p}{1-p}\right)^{c}}$$

$$(5) \qquad = \left(1-\frac{1}{N}\right)\frac{\binom{m}{b\,c}}{\binom{m}{b}\binom{m}{c}}\left(\frac{1-2p}{(1-p)(1-p)}\right)^m\left(\frac{1-p}{1-2p}\right)^{b+c}$$

$$= \left(1-\frac{1}{N}\right)\frac{(m-c)^{(b)}}{m^{(b)}}\left(1-\left(\frac{p}{1-p}\right)^2\right)^m\left(\frac{1-p}{1-2p}\right)^{b+c}$$

where $u^{(b)} = u(u-1)\cdots(u-b+1)$ denotes a descending factorial with b factors.

Notice that, if the n_b were independently distributed in Poisson distributions, the second moments would be given by the same formula with $\Phi(b, c) = 1$, while if they were distributed like a multinomial sample from an infinite population the second moments would be given by the same formula with $\Phi(b, c) = 1 - \frac{1}{N}$.

For small p, we have

$$\Phi(b, c) \approx \left(1-\frac{1}{N}\right)\frac{(m-c)^{(b)}}{m^{(b)}},$$

and if m is large compared to b and c, this approaches the multinomial value

$$\Phi(b, c) \approx \left(1-\frac{1}{N}\right).$$

13. Variances and covariances. The variances and covariances are given by

$$\begin{aligned}\text{Variance } (n_b) &= E(n_b n_b) - E(n_b)E(n_b)\\ &= E(n_b)(1-(1-\Phi(b, b))E(n_b)),\end{aligned}$$

and

$$\text{Covariance } (n_b, n_c) = -(1-\Phi(b, c))E(n_b)E(n_c).$$

Thus the covariance of n_b and n_c will vanish when, and only when $\Phi(b, c) = 1$.

Let us suppose $pN = \frac{1}{\beta}$, with p small and m and N large, and see if $\Phi(b, c)$ can be unity. Since a preliminary calculation shows it to be reasonable, let us put $m = \gamma N$. Then

$$\Phi(b, c) \approx (1-\beta p)\frac{(\gamma N - c)^{(b)}}{(\gamma N)^{(b)}}(1-p^2)^{\gamma N}(1+p)^{b+c}.$$

An easy calculation shows that the ratio of descending factorials is nearly

$$e^{-bc/\gamma N} = e^{(-bc\beta/\gamma)p},$$

making further natural approximations,

$$\ln \Phi(b, c) \approx -\beta p - \frac{bc\beta}{\gamma} p - \gamma N p^2 + (b + c)p$$

and this may be written

$$\ln \Phi(b, c) \approx -\frac{\beta p}{4\gamma} \left(\left(2\frac{\gamma}{\beta} - b - c + \beta \right)^2 + 4\beta c - (b - \beta - c)^2 \right),$$

and this vanishes for real γ when and only when $| b - \beta - c | \geq \sqrt{4\beta c}$. This, then, is the condition on b and c which permits the existence of two ratios of m to N so that for either ratio and large N there will be no correlation between n_b and n_c.

14. Higher moments. To deal with the third moments, we need $E(x_{iq} y_{jq} z_{kq})$, which is easily seen to behave as follows:

Relation of ijk	*number of occurrences*	*Expectation of* $x_{iq} y_{jq} z_{kq}$
$i = j = k$	N	$1 - p + pxyz$
$i = j \neq k$	$N(N - 1)$	$1 - 2p + pxy + pz$
$i = k \neq j$	$N(N - 1)$	$1 - 2p + pxz + py$
$j = k \neq i$	$N(N - 1)$	$1 - 2p + pyz + px$
different	$N(N - 1)(N - 2)$	$1 - 3p + px + py + pz$

Thus we have

$$\begin{aligned}\Sigma_{bcd} x^b y^c z^d E(n_b n_c n_d) &= N(1 - p + pxyz)^m + N(N - 1)(1 - 2p + pxy + pz)^m \\ &+ N(N - 1)(1 - 2p + pxz + py)^m + N(N - 1)(1 - 2p + pyz + px)^m \\ &+ N(N - 1)(N - 2)(1 - 3p + px + py + pz)^m\end{aligned}$$

from which we can calculate all third moments.

In general if ϵ is a decomposition of the product $xyz \cdots w$ into α monomials $u_1, u_2, \cdots, u_\alpha$, where order is disregarded (for example: $xyz = (xz)y = (zx)y = y(zx) = y(xz)$ is a single decomposition with $\alpha = 2$, $u_1 = xz$, $u_2 = y$), then the generating function becomes

$$\Sigma_\epsilon N^{(\alpha)}(1 + (u_1 + u_2 + \cdots + u_\alpha - \alpha)p)^m.$$

15. Poisson balls and equal boxes. To reach a Poisson distribution we let $m \to \infty$ and $p \to 0$ so that $mNp = tN$, where t is the *average number of balls per box* in the Poisson distribution.

Since

$$p^b \binom{m}{b} \to \frac{t^b}{b!}$$

under these conditions, (3) becomes

$$(6) \qquad E(n_b) = N \frac{t^b}{b!} e^{-t}$$

and from (5) it follows that the limit of $\Phi(b, c)$ is $\left(1 - \frac{1}{N}\right)$ so that

$$(7) \qquad E(n_b n_c) = N(N-1) \frac{t^{b+c}}{b!\,c!} e^{-2t} + \delta(b, c) N \frac{t^b}{b!} e^{-t},$$

and hence

$$(8) \qquad \text{Variance } (n_b) = N \left(\frac{t^b}{b!} e^{-t}\right)\left(1 - \frac{t^b}{b!} e^{-t}\right),$$

$$(9) \qquad \text{Covariance } (n_b, n_c) = -N \left(\frac{t^b}{b!} e^{-t}\right)\left(\frac{t^c}{c!} e^{-t}\right).$$

Notice that these are the moments of the numbers of objects of types $b, c, \cdots$, in a random sample of N from an infinite population where the fraction of b's is $t^b e^{-t}/b!$, just as it should be.

16. Fixed or binomial balls and varied boxes. We now consider the case where the chance of any ball entering the ith box is p_i. We shall again not restrict ourselves to the case $\Sigma_i p_i = 1$.

The expectation of x_{iq} is immediately seen to be $(1 + p_i(x - 1)) = (1 - p_i + p_i x)$, so that the generating function is

$$f(x) = \Sigma_i (1 - p_i + p_i x)^m$$

and the expectation of n_b is

$$(10) \qquad E(n_b) = \binom{m}{b} \Sigma_i (1 - p_i)^{m-b} p_i^b = \binom{m}{b} \Sigma_i (1 - p_i)^m \left(\frac{p_i}{1 - p_i}\right)^b.$$

Following Stevens [4] with a slight modification, let us set $p_i = p(1 + \lambda_i)$, where p is the average of the p_i, so that $\Sigma_i \lambda_i = 0$. Then

$$(1 - p_i) = (1 - p(1 + \lambda_i)) = (1 - p)\left(1 - \frac{p\lambda_i}{1 - p}\right),$$

so that

$$\Sigma_i (1 - p_i)^{m-b} p_i^b = (1 - p)^{m-b} p^b \Sigma_i \left(1 - \frac{p\lambda_i}{1 - p}\right)^{m-b} (1 + \lambda_i)^b.$$

Expanding the summand, we find

$$1+\left\{-\frac{(n-b)p}{1-p}+b\right\}\lambda_i$$

$$+\left\{\frac{(m-b)(m-b-1)p^2}{2(1-p)^2}-\frac{(m-b)bp}{1-p}+\frac{b(b-1)}{2}\right\}\lambda_i^2+0(\lambda_i^3).$$

Hence, setting $\Sigma_i\lambda_i^2 = \Lambda$ (notice this is not the same as Stevens' Λ!), we have

$$E(n_b)=\binom{m}{b}(1-p)^{m-b}p^b$$

$$\left\{N+\tfrac{1}{2}\Lambda\left(\frac{m-b}{m-b-1}\,\frac{(p(m-1)-b)^2}{(1-p)^2}-\frac{b(m-1)}{m-b-1}\right)\right\}+O(\Sigma_i\lambda_i^3).$$

The expectation for all $p_i = p$ has been modified by multiplication by

$$1+\frac{\Lambda}{2N}\left\{\frac{m-b}{m-b-1}\,\frac{(p(m-1)-b)^2}{(1-p)^2}-\frac{b(m-1)}{m-b-1}\right\} \tag{11}$$

plus terms of higher order. For large N and consequently small p the quantity in braces is nearly

$$b\left(b-\frac{m}{m-b}\right)$$

and more roughly is approximately b^2. Similarly, the expectations of second moments are

$$E(n_b n_c)=\binom{m}{b\,c}\sum_{i\neq j}(1-p_i-p_j)^{m-b-c}p_i^b p_j^c+\delta(b,c)\binom{m}{b}\sum\nolimits_i(1-p_i)^{m-b}p_i^b,$$

whence

$$\Phi(b,c)=\frac{\binom{m}{b\,c}\sum_{i\neq j}(1-p_i-p_j)^{m-b-c}p_i^b p_j^c}{\binom{m}{b}\binom{m}{c}\sum_i(1-p_i)^{m-b}p_i^b\sum_j(1-p_j)^{m-c}p_j^c}. \tag{12}$$

Making the same sort of expansion yields

$$\Phi(b,c)\approx\left(1-\frac{1}{N}\right)\frac{(m-c)^{(b)}}{m^{(b)}}\left(1-\frac{p^2}{(1-p)^2}\right)^m\left(\frac{1-p}{1-2p}\right)^{b+c}\left(1+\frac{\Lambda\psi}{2N}\right) \tag{13}$$

where terms in $\Sigma\lambda_i^3$ have been neglected (note that

$$\sum_{i\neq j}\lambda_i\lambda_j=-\sum_i\lambda_i^2=-\Lambda\cdot),$$

and where

$$\psi=\left\{\frac{m-b-c}{m-b-c-1}\,\frac{N-2}{N-1}(1-2p)^{-2}-\frac{m-b}{m-b-1}(1-p)^{-2}\right\}$$

$$\cdot\{p(m-1)-b\}^2$$

$$+\left\{\frac{m-b-c}{m-b-c-1}\,\frac{N-2}{N-1}\,(1-2p)^{-2}-\frac{m-c}{m-c-1}\,(1-p)^{-2}\right\}$$
$$\cdot\{p(m-1)-c\}^2$$
$$+\frac{1}{2}\,\frac{m-b-c}{m-b-c-1}\left\{\frac{N}{N-1}-\frac{N-2}{N-1}\,(1-2p)^{-2}\right\}(b-c)^2$$
$$+\frac{1}{m-b-c-1}\left\{\frac{2bc}{N-1}-\frac{b^2c}{m-b-1}-\frac{c^2b}{m-c-1}\right\}.$$

This can be reduced to

$$\psi = 2bc\left(2p-\frac{1}{N}\right)+\mathrm{O}(p^2)+\mathrm{O}\left(\frac{b}{N}\right),$$

and for $p = 1/N + \mathrm{O}(p^2) + \mathrm{O}\left(\frac{b}{N}\right)$.

$$\psi = 2pbc + \mathrm{O}(p^2).$$

17. Poisson balls and varied boxes. To reach the Poisson limit, we let $m \to \infty$ and $p_i \to 0$ so that $mp_i = t_i$. The generating function for first moments becomes

$$f(x) = \Sigma_i e^{-t_i + t_i x}$$

and the expectation of n_b is

$$E(n_b) = \Sigma_i \frac{t_i^b}{b!}\, e^{-t_i}. \tag{15}$$

If we set $t_i = t(1 + \Lambda_i)$, this becomes

$$E(n_b) = \frac{t^b}{b!}\, e^{-t}\, \Sigma_i (1+\lambda_i)^b e^{-t\lambda_i}$$

The summand expands in the form

$$\left(1 + b\lambda_i + \frac{b(b-1)}{2}\lambda_i^2 + \frac{b(b-1)(b-2)}{6}\lambda_i^3 + \cdots\right)$$
$$\times\left(1 - t\lambda_i + \frac{t^2}{2}\lambda_i^2 - \frac{t^3}{6}\lambda_i^3 + \cdots\right)$$
$$= 1 + (b-t)\lambda_i + \left(\frac{b(b-1)}{2} - bt + \frac{t^2}{2}\right)\lambda_i^2 + \cdots.$$

If t is chosen as the average of the t_i so that $\Sigma\lambda_i = 0$, the sum becomes

$$N + \left(\frac{(b-t)^2 - b}{2}\right)\Sigma\lambda_i^2 + \left(\frac{(b-t)^3}{6} - \frac{3b-2}{6} + \frac{bt}{2}\right)\Sigma\lambda_i^3 + \cdots.$$

Again setting $\Sigma\lambda_i^2 = \Lambda$ we have

$$E(n_b) \approx \frac{t^b}{b!}\, e^{-t}\left(N + \left(\frac{(b-t)^2 - b}{2}\right)\Lambda\right) \tag{16}$$

which can be written

$$E(n_b) \approx N \frac{t^b}{b!} e^{-t}\left(1 + \frac{\Lambda}{2N}\left((b-t)^2 - b\right)\right).$$

The generating function for the second moments is

$$f(x)f(y) = \Sigma_{ij} e^{-t_i + t_i x - t_j + t_j y}$$

so that the expectation of $n_b n_c$ is

$$E(n_b n_c) = \sum_{i \neq j} \frac{t_i^b t_j^c e^{-t_i - t_j}}{b!c!} + \delta(b, c) \sum_i \frac{t_i^b}{b!} e^{-t_i} \tag{17}$$

which becomes

$$E(n_b n_c) = \frac{t^{bc}}{b!c!} e^{-2t} \sum_{i \neq j} (1 + \lambda_i)^b (1 + \lambda_j)^c e^{-t\lambda_i - t\lambda_j} + \delta(b, c) E(n_b),$$

whence we can derive

$$\Phi(b, c) \approx 1 - \frac{1}{N} - \frac{\Lambda}{2N^2}(b - t)(c - t). \tag{18}$$

Thus

$$\text{Variance } (n_b) \approx E(n_b) - \frac{1}{N}\left\{1 + \frac{\Lambda}{2N}(b - t)^2\right\}(E(n_b))^2, \tag{19}$$

$$\text{Covariance } (n_b n_c) \approx -\frac{1}{N}\left(1 + \frac{\Lambda}{2N}(b - t)(c - t)\right) E(n_b)E(n_c). \tag{20}$$

18. Boxes in a systematic square. Another case which it may be worthwhile to write down arises when the boxes are systematically "rotated" under "spouts" of different probability. That is, the number of balls m is a multiple of the number of boxes N, and the probability of the qth ball entering the ith box depends on the value of $q - i$ taken modulo N. An example for $N = 3$ and $m = 6$ follows:

Probabilities of entry

Box	Ball 1	2	3	4	5	6
1	p_0	p_1	p_2	p_0	p_1	p_2
2	p_2	p_0	p_1	p_2	p_0	p_1
3	p_1	p_2	p_0	p_1	p_2	p_0

If $m = kN$ and the subscript r runs through $0, 1, 2, \cdots, N - 1$, then the expectation of $f(x)$ becomes

$$\Sigma_i \Sigma_q E(x_{iq}) = N\{\Pi_r(1 - p_r + p_r x)\}^m.$$

Thus first moments, and by proceeding similarly higher moments, are available for this case also.

PART III

THE POISSON CASE

19. The Poisson case with equal boxes. The Poisson case is obtained in the limit as $m \to \infty$ and $p \to 0$ with $pm = t$. We wish to show that, in the limit, the number of balls in the different boxes are independent. Let $k_1, k_2, \cdots, k_N$ be the number of balls in the first, second, $\cdots$, Nth box, respectively. Then the distribution of the k's is given by, where we write $k = k_1 + k_2 + \cdots + k_N$,

$$\frac{m^{(k)}}{k_1!k_2! \cdots k_N!} p^k(1 - Np)^{m-k} = \frac{m^{(k)}}{m^k} \frac{(1 - Np)^{m-k}}{e^{-Nmp}} \Pi_i \frac{(mp)^{k_i}e^{-mp}}{k_i!}$$

Now the first two fractions clearly approach unity in the limit, and the independence is proved.

Since the number of balls in each box has an independent Poisson distribution, the distribution of the numbers of boxes each with exactly b balls is that of a random sample of N from an infinite population—namely it is a multivariate distribution with probabilities

$$\frac{(mp)^b e^{-mp}}{b!} .$$

REFERENCES

[1] H. H. Newcombe, "Delayed phenotypic expression of spontaneous mutations in *Escherichia coli*," *Genetics*, *Vol. 33* (1948), p. 447–476.

[2] I. Opatowski, "Chain processes and their biophysical applications: Part I. General Theory," *Bulletin of Mathematical Biophysics*, *Vol. 7* (1945), p. 161–180.

[3] V. Romanovsky, "Su due problemi di distribuzione casuale," *Giornale dell'Istituto Italiano degli Attuari*, *Vol. 5* (1934), p. 196–218.

[4] W. L. Stevens, "Significance of grouping," *Annals of Eugenics*, *Vol. 8* (1937–1938), p. 57–59.

THE POWER OF THE CLASSICAL TESTS ASSOCIATED WITH THE NORMAL DISTRIBUTION

By J. Wolfowitz

Columbia University

Summary. The present paper is concerned with the power function of the classical tests associated with the normal distribution. Proofs of Hsu, Simaika, and Wald are simplified in a general manner applicable to other tests involving the normal distribution. The set theoretic structure of several tests is characterized. A simple proof of the stringency of the classical test of a linear hypothesis is given.

1. Introduction. The present paper is concerned with the optimum properties, from the power function viewpoint, of the classical tests associated with the normal distribution. In 1941 Hsu [2] proved the result stated in Section 2 below, which is concerned with the general linear hypothesis (in this connection his paper [1] of 1938 will be of interest). Also in 1941 Simaika [3] proved similar results for the tests based on the multiple correlation coefficient and Hotelling's generalization of Student's t. In 1942, Wald [4] gave a generalization of Hsu's result.

In the present paper we give short and simple proofs of almost all these results, and a simple proof of the stringency property of the analysis of variance (Section 5). These proofs rest on theorems which characterize the set theoretic structure of the tests. Thus, while the proofs of Hsu, Simaika and Wald are rather elaborate and each problem is essentially attacked *de novo*, the methods of the present paper are in effect applicable to the classical tests based on the normal distribution. For these tests it will not be difficult to demonstrate the analogues of Theorems 1 and 3, and of the results of Hsu, Simaika, and Wald. In the present paper we first treat the general linear hypothesis, because it is the simplest problem, its solution is easiest to describe, and it admits Wald's integration theorem. Multivariate analogues of the latter are rather artificial and not as simple. We then discuss the problem of the multiple correlation coefficient, because it seems to be more difficult than that of Hotelling's T and indeed, to include all the essential multivariate difficulties. Theorems 6 and 7 are the analogues of 1 and 3, respectively, while Theorem 9 describes the essential property of the power function which is of interest to us. In other multivariate problems one will prove the analogues of Theorems 6, 7 and 9. A generally inclusive formulation is no doubt possible. Theorems 5 and 9 are slightly more general than the theorems of Hsu and Simaika.

Many of the statements below may be not valid on exceptional sets of measure zero. Usually this is so stated, but sometimes, for reasons of brevity or to avoid repetition, this qualification may be omitted. The reader will have no difficulty supplying it wherever necessary.

The author is indebted to Erich L. Lehmann of the University of California, who carefully read a first version of this paper. Theorem 4 below was arrived at independently by Professor Lehmann, with a somewhat different proof.

2. The general linear hypothesis. In canonical form the general linear hypothesis may be stated as follows: The chance variables

$$X_1, X_2, \cdots, X_{k+l}$$

have at $x_1, \cdots, x_{k+l}$, the density function

$$(\sqrt{2\pi}\,\sigma)^{-(k+l)} \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_1^k (x_i - \eta_i)^2 + \sum_{k+1}^{k+l} x_i^2\right\}\right] = f(\eta, \sigma) \tag{2.1}$$

with $\sigma, \eta_1, \cdots, \eta_k$ all unknown.

Let η be the vector $(\eta_1, \cdots, \eta_k)$. The null hypothesis H_0 states that

$$\eta_1 = \cdots = \eta_k = 0$$

and is to be tested with constant size $\alpha < 1$ (identically in σ).

Let D be any admissible critical region for testing H_0. If A is any event let $P\{A \mid \eta, \sigma\}$ denote the probability of A when η and σ are the parameters of (2.1). We have then

$$P\{D \mid 0, \sigma\} = \alpha$$

identically in σ, where 0 is the vector with k components all of which are zero. We now prove a property which characterizes all D. This theorem is due to Neyman and Pearson [12], and is given here only for completeness.

THEOREM 1. *The fraction of the surface area of the sphere*

$$\sum_1^{k+l} x_i^2 = c^2$$

which lies in D is α for almost all c.

PROOF. Let a be any positive integer, h a positive parameter, and $\psi(y)$ a measurable function of y defined for $y > 0$ and such that $0 \le \psi(y) \le 1$. In view of the distribution of ΣX_i^2, it will be enough to prove that, if

$$\frac{h^{a+1}}{\Gamma(a+1)} \int_0^\infty \psi(y) y^a e^{-hy}\, dy = \alpha$$

identically for all positive h, that then

$$\psi(y) = \alpha \text{ for almost all } y.$$

Write

$$\frac{1}{\alpha\Gamma(a+1)} \int_0^\infty \psi(y) y^a e^{-hy}\, dy = h^{-(a+1)}. \tag{2.2}$$

Differentiating both members k times with respect to h and then setting $h = 1$

we obtain the following result. The function

$$\frac{1}{\alpha\Gamma(a+1)}\,\psi(y)y^{a}e^{-y}$$

is a density function with kth moment

$$\mu_k = (a+1)(a+2)\cdots(a+k).$$

The moments μ_k are the moments of the density function

$$\frac{1}{\Gamma(a+1)}\,y^{a}e^{-y}.$$

They satisfy the Carleman criterion [5, p. 19, Th 1.10], and hence no essentially different distribution can have these moments. This proves the desired result.

Theorem 2 (Wald). *Among all tests of the general linear hypothesis the analysis of variance test has the property that, for all positive d, the integral of its power on the surface $\eta^2 = d^2$ is a maximum.*

Proof. Let c be any positive number. We have only to show that if we allocate to the critical region D of the test the fraction α of the surface area of the sphere

$$\sum_{1}^{k+l} x_i^2 = c^2 \tag{2.3}$$

for which

$$C = \frac{\sum_{1}^{k} x_i^2}{\sum_{k+1}^{k+l} x_i^2}$$

is as large as possible and that if we do this for all c, the desired maximum of the integral of the power will be achieved. If C is as large as possible so is

$$\frac{\sum_{1}^{k} x_i^2}{\sum_{1}^{k+l} x_i^2} = \frac{\sum_{1}^{k} x_i^2}{c^2}.$$

Let $a_1, \cdots, a_{k+l}$ be any point on the sphere (2.3). Let db be the differential of area on the surface $\eta^2 = d^2$. Then

$$\int_{\eta^2=d^2}\cdots\int f(\eta, \sigma)\, db = (\sqrt{2\pi}\,\sigma)^{-(k+l)} \exp\left\{-\frac{(c^2+d^2)}{2\sigma^2}\right\} \cdot \int_{\eta^2=d^2}\cdots\int \exp\left\{\frac{(\eta)'z}{\sigma^2}\right\} db, \tag{2.4}$$

where z is the vector $(a_1, \cdots, a_k)$ and $(\eta)'z$ is the scalar product of the two vectors. This last integral is easily seen to depend only upon $|z|$ and to be monotonically increasing in $|z|$. This proves the theorem.

COROLLARY (Hsu). *Among all tests of the general linear hypothesis whose power is a function of η^2 only, the analysis of variance is the most powerful.*

3. The set theoretic structure of tests whose power is a function only of η^2/σ^2. Wald's result (Theorem 2) cannot always be extended, in its simple form, to tests involving the multivariate normal distribution, but this can be done with Hsu's theorem (corollary to Theorem 2). In order to see what is involved we shall investigate the set theoretic structure of tests of the general linear hypothesis whose power is a function only of η^2/σ^2.

Let $q(x_1, \cdots, x_k)$ be the set of points in the region D whose first k coordinates are $x_1, \cdots, x_k$. Let $A(x_1, \cdots, x_k, \sigma)$ be the integral of

$$(2\pi\sigma^2)^{-(l/2)} \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_{j=1}^{l} x_{k+j}^2\right\}\right]$$

with respect to $x_{k+1}, \cdots, x_{k+l}$, taken over $q(x_1, \cdots x_k)$. We first prove the following:

LEMMA. *Suppose the power of D is a function only of η^2/σ^2. Then for two points*

$$x_1, \cdots, x_k$$

and

$$x_1', \cdots, x_k'$$

such that

$$\sum_1^k x_i^2 = \sum_1^k x_i'^2 \tag{3.1}$$

we have

$$A(x_1, \cdots, x_k, \sigma) = A(x_1', \cdots, x_k', \sigma) \tag{3.2}$$

identically in σ, with the exception of a set of measure zero.

PROOF. Suppose the statement is false. Then under some orthogonal transformation T of $x_1, \cdots, x_k$ the region D would go over into a region D^* with the following property: Let $A^*(x_1, \cdots, x_k, \sigma)$ have the same definition for the region D^* as $A(x_1, \cdots, x_k, \sigma)$ has for D. Then on a set of positive measure[1] we would have

$$A(x_1, \cdots, x_k, \sigma) \neq A^*(x_1, \cdots, x_k, \sigma). \tag{3.3}$$

We shall now show that (3.3) results in a contradiction. We have

$$P\{D \mid \eta, \sigma\} = P\{D^* \mid T\eta, \sigma\} \tag{3.4}$$

identically in η. By the property of the region D, therefore, we have

$$P\{D \mid \eta, \sigma\} = P\{D \mid T^{-1}\eta, \sigma\}$$

[1] The situation here is similar to that described in footnote 3.

and hence

$$P\{D \mid \eta, \sigma\} = P\{D^* \mid \eta, \sigma\} \tag{3.5}$$

identically in η. Thus we obtain

$$\begin{aligned} &\int (2\pi\sigma^2)^{-(k/2)} A(x_1, \cdots, x_k, \sigma) \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_1^k (x_i - \eta_i)^2\right\}\right] dx_1 \cdots dx_k \\ &\equiv \int (2\pi\sigma^2)^{-(k/2)} A^*(x_1, \cdots, x_k, \sigma) \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_1^k (x_i - \eta_i)^2\right\}\right] dx_1 \cdots dx_k \end{aligned} \tag{3.6}$$

with the integrations taking place over the entire space. Differentiating both members with respect to the components of η and setting $\eta = 0$, we obtain that the two density functions (for fixed σ)

$$(2\pi\sigma^2)^{-(k/2)}\alpha^{-1}A(x_1, \cdots, x_k, \sigma) \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_1^k x_i^2\right\}\right]$$

and

$$(2\pi\sigma^2)^{-(k/2)}\alpha^{-1}A^*(x_1, \cdots, x_k, \sigma) \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_1^k x_i^2\right\}\right]$$

have identical moments. We shall now argue that these moments satisfy the conditions of Cramér and Wold [7, Th. 2], so that the two density functions are essentially the same, in contradiction to (3.3). The Cramér-Wold theorem states the following: *Let $Y_1, \cdots, Y_k$ be k chance variables with a joint distribution function, and write*

$$\lambda_{2n} = \sum_{i=1}^k EY_i^{2n}.$$

Then the divergence of the series

$$\sum_{n=1}^\infty \lambda_{2n}^{-(1/2n)}$$

is sufficient to ensure that there exists essentially only one distribution which has these moments. We notice that the factor $1/\alpha$ of course makes no difference. If we set $A(x_1, \cdots, x_k, \sigma)$ and $A^*(x_1, \cdots, x_k, \sigma)$ both identically unity and consider the resulting moments which enter into the λ_{2n}, we see that these moments satisfy the Cramér-Wold condition. Now A and A^* are ≤ 1. Thus, using the true value of A can serve only to increase the value of $\lambda_{2n}^{-(1/2n)}$, so that the series will diverge a fortiori. This proves the lemma.

The following theorem helps to describe the set theoretic structure of tests whose power is a function only of $\lambda = \eta^2/\sigma^2$:

THEOREM 3. *Let D be a test whose power is a function only of λ. Let u be any positive number, and $D(x_1, \cdots, x_k, u)$ be the fraction of the "area" of the sphere $\Sigma_{j=1}^l x_{k+j}^2 = u^2$ occupied by points which are in D and whose first k coordinates are $x_1, \cdots, x_k$. If*

$$\sum_1^k x_j^2 = \sum_1^k x_j'^2 \tag{3.7}$$

then, except on a set of measure zero,

$$D(x_1, \cdots, x_k, u) = D(x_1', \cdots, x_k', u). \tag{3.8}$$

PROOF. We shall show that, if the power of D is a function only of λ, the failure of (3.7) to imply (3.8) would contradict the preceding lemma. Suppose then that (3.8) is not true on a set of positive measure. Under some orthogonal transformation on $x_1, \cdots, x_k$ we obtain[2] a function $D^*(x_1, \cdots, x_k, u)$ which differs from $D(x_1, \cdots, x_k, u)$ on a set of positive measure and such that, for almost every $x_1, \cdots, x_k$,

$$A(x_1, \cdots, x_k, \sigma) = K \int_0^\infty D(x_1, \cdots, x_k, u)\sigma^{-l} u^{l-1} e^{(-u^2)/2\sigma^2}\, du$$

$$= K \int_0^\infty D^*(x_1, \cdots, x_k, u)\sigma^{-l} u^{l-1} e^{(-u^2)/2\sigma^2}\, du$$

identically in σ, where K is a suitable constant of no interest to us. Multiplying by σ^l, differentiating repeatedly under the integral sign with respect to σ, and setting $\sigma = 1$, we obtain the result that the two density functions in u,

$$\frac{KD(x_1, \cdots, x_k, u)}{A(x_1, \cdots, x_k, 1)} u^{l-1} e^{(-u^2)/2}$$

and

$$\frac{KD^*(x_1, \cdots, x_k, u)}{A(x_1, \cdots, x_k, 1)} u^{l-1} e^{(-u^2)/2}$$

are identical except perhaps on a set of measure zero. This contradiction proves the theorem.

THEOREM 4. *A necessary and sufficient condition that the power of D be a function of λ only, is that, with the usual exception of a set of measure zero, $D(x_1, \cdots, x_k, u)$ be a function only of*

$$\frac{\sum_1^k x_i^2}{u^2}.$$

The proof of this theorem is not essentially different from that of the preceding theorem, and we shall therefore sketch it only briefly. Let Z be a transformation on $(x_1, \cdots, x_k, u) = (x, u)$ which consists of a rotation of the vector x, followed by a multiplication of u and the components of x by a positive constant c. If $D(x, u)$ is not a function of $\Sigma_1^k x_i^2/u^2$ alone, then, just as before[3], we can use some

[2] See footnote 1.

[3] This statement implies that a function of $x_1, \cdots, x_k, u$, which is invariant to within sets of measure zero under all transformations Z (the exceptional set may depend on the

transformation Z to give us a function $D^*(x, u)$ such that

$$D(x, u) \neq D^*(x, u)$$

on a set of positive measure, while

$$ED(x, u) = ED^*(x, u)$$

identically in η, σ. This yields a contradiction in the usual manner and proves the necessity of the condition.

To prove sufficiency, write $D(x, u) = \nu(\Sigma x_i^2/u^2) = \nu(v)$. Let $\gamma(v, \eta, \sigma)$ be the density function of v. Then

$$P\{D \mid \eta, \sigma\} = \int_0^\infty \nu(v)\gamma(v, \eta, \sigma)\, dv.$$

By hypothesis, $\nu(v)$ is a function only of v. We know [9, p. 140, eq. 101] that $\gamma(v, \eta, \sigma)$ is a function only of v and λ. Hence $P\{D \mid \eta, \sigma\}$ is a function only of λ. This completes the proof of the theorem.

THEOREM 5. *Among all tests of the general linear hypothesis which have the properties described in the conclusions of Theorems* 1 *and* 3, *the classical analysis of variance test is the most powerful.*

We shall omit the proof of this theorem, which is very similar to that of the more difficult Theorem 9 below.

Theorem 4 above shows that there exist regions D which satisfy the conclusions of Theorems 1 and 3 and such that $P\{D \mid \eta, \sigma\}$ is not a function of λ alone. It follows that the content of Theorem 5 is greater than that of Hsu's theorem (Corollary to Theorem 2).

It is instructive to note that Hsu's theorem follows almost immediately from Theorem 4 and the form of $\gamma(v, \lambda)$. For let λ be fixed but arbitrary. One verifies immediately from the form of $\gamma(v, \lambda)$ that

$$\frac{\gamma(v, \lambda)}{\gamma(v, 0)}$$

is, for fixed λ, a monotonically increasing function of v. This, by Neyman's lemma, immediately proves Hsu's result.

4. The multiple correlation coefficient. We shall now apply our methods to a multivariate test. For typographic ease we shall conduct the discussion for the

---

transformation), is a function of $\frac{\Sigma x_i^2}{u^2}$, except on a set of measure zero. This statement would be completely trivial were it not for the exceptional sets; in any case it must be well known to set theorists. The author constructed an unnecessarily long proof of it, and believes that a more expeditious proof can be constructed using the ideas of [11, page 91, Theorem 11.1, and page 318, p. 7]. Professor C. M. Stein of the University of California has informed the author that this result is a special case of one established by himself and G. H. Hunt in a forthcoming paper. For these reasons the proof is omitted. (See also [13, page 27, Lemma 9.1].)

case of three variates, but the reader will observe that the procedure is really perfectly general.

The chance variables $\{Y_{ij}\}$, $i = 1, 2, 3$, $j = 1, \cdots, n$, have the density function

$$(4.1) \qquad g(B) = (2\pi)^{(-3n)/2}(|B|)^{n/2} \exp\left\{-\tfrac{1}{2}\sum_{j=1}^{n}\sum_{i,l=1}^{3} b_{il}y_{ij}y_{lj}\right\}$$

where 1) $B = \{b_{il}\}$ is a positive definite (symmetric) 3×3 matrix, 2) y_{ij} is the value assumed by Y_{ij}. The null hypothesis H_0 asserts that a given multiple correlation coefficient is zero, say that of Y_1 with Y_2 and Y_3, i.e.,

$$(4.2) \qquad b_{12} = b_{21} = b_{13} = b_{31} = 0.$$

The test is to be made on the level of significance α, i.e., if B_0 is any matrix which satisfies (4.2), and if G is a critical region for testing H_0, then

$$(4.3) \qquad P\{G \mid B_0\} = \alpha$$

where the symbol in the left member means the probability of G according to $g(B_0)$.

Write

$$ns_{ij} = \sum_{k=1}^{n} y_{ik}y_{jk}$$

$$S = \begin{Bmatrix} s_{22} & s_{23} \\ s_{32} & s_{33} \end{Bmatrix}.$$

Let $M(c_{11}, C)$ be the manifold in the $3n$-space of

$$y_{11}, \cdots, y_{ik}, \cdots, y_{3n}$$

where $s_{11} = c_{11}$, $S = C$. First we prove the following:

THEOREM 6. *Any region G which satisfies* (4.3) *must have the property that the fraction of the area of $M(c_{11}, C)$ which lies in G is α, for any positive c_{11} and any positive definite 2×2 matrix $C = \{c_{ij}\}$*. (We remind the reader that exceptional sets of measure zero are not precluded).

PROOF. Let $\psi(c_{11}, C)$ be the fraction of the area of $M(c_{11}, C)$ in G. Recall equation (4.3) and the fact that $s_{11}, s_{22}, s_{23}, s_{33}$ are sufficient statistics for the elements of B_0. On the manifold $M(c_{11}, C)$ the conditional density is uniform. Employing Wishart's distribution [6] we conclude that

$$(4.4) \qquad K' \int \psi(s_{11}, S) \mid B_0 \mid^{N} \mid S \mid^{(n-3)/2} s_{11}^{(n-2)/2}$$

$$\cdot \exp\left[-\frac{n}{2}\{b_{11}s_{11} + b_{22}s_{22} + 2b_{23}s_{23} + b_{33}s_{33}\}\right] ds_{11}\, ds_{22}\, ds_{23}\, ds_{33} \equiv \alpha$$

where K' is a suitable constant which need not concern us. Here the symbol

"≡" means identically in b_{11}, b_{22}, b_{23}, b_{33}, provided only that $b_{11} > 0$, $b_{22} > 0$ $b_{22}b_{33} - b_{23}^2 > 0$. Of course s_{11} is distributed independently of s_{22}, s_{23}, s_{33}. Proceeding as in section 2, we can, by differentiation with respect to the b's, obtain all the moments of the s_{ij}'s. Now let the b's take any admissible constant values. The moments of the s_{ij}'s are then seen to satisfy the criterion of Cramér and Wold [7, Th. 2], and consequently essentially uniquely determine the distribution of the s_{ij}. The desired conclusion follows as before.

The six parameters which uniquely determine the trivariate normal distribution (of Y_1, Y_2, Y_3) with zero means may be taken to be the following:

1) The covariance matrix $\{\sigma_{ij}\}$, $i, j = 2, 3$, of Y_2 and Y_3.

2) The partial regression coefficients β_2, β_3, of Y_1 on Y_2 and Y_3. These are defined as follows: Let $E(Y_1 \mid Y_2 = y_2, Y_3 = y_3)$ denote the conditional expected value of Y_1, given $Y_2 = y_2$, $Y_3 = y_3$. Then

$$E(Y_1 \mid Y_2 = y_2, Y_3 = y_3) = \beta_2 y_2 + \beta_3 y_3 .$$

3) The conditional variance ω^2 of Y_1, given $Y_2 = y_2$, $Y_3 = y_3$.

The population multiple correlation coefficient $\bar{R}$ of Y_1 with Y_2 and Y_3 is then defined by

$$\frac{\bar{R}^2\omega^2}{(1 - \bar{R}^2)} = \beta_2^2\sigma_{22} + 2\beta_2\beta_3\sigma_{23} + \beta_3^2\sigma_{33}.$$

The six parameters above may be chosen arbitrarily, provided only that $\{\sigma_{ij}\}$ is positive definite. $\bar{R}$ and ω are, by definition, non-negative.

Let y_i be the column vector $y_{i1}, \cdots, y_{in}$; let y_i' be its transpose, and let y denote the point $y_{11}, y_{12}, \cdots, y_{1n}, y_{21}, \cdots, y_{3n}$ in $3n$-space. Let $z(y) = z(y_1, y_2, y_3)$ be the component of y_1 in the plane of y_2 and y_3; let $r = | z(y) |$ and θ the angle between z and y_2, measured positively say in the direction of y_3. Finally let h be the absolute value of the vector $y_1 - z(y_1, y_2, y_3)$.

We intend now to investigate the set theoretic structure of tests whose power is a function only of $\bar{R}$, and for this purpose prove the following:

THEOREM 7. *Let H be a region whose power is a function only of $\bar{R}$. Let $V(h, r, \theta, s_{22}, s_{23}, s_{33})$ be the fraction of the "volume" of the manifold on which $h, r, \theta, s_{22}, s_{23}, s_{33}$ are fixed which is contained in H. With the usual exception of a set of measure zero, for fixed $h, r, s_{22}, s_{23}, s_{33}$, the quantity V above is constant for all θ.*

Later, after this theorem is proved, we shall write V without exhibiting θ. This procedure is justified by Theorem 7.

PROOF. Suppose the theorem false, and proceed as in Theorem 3. A suitable[4] rotation of the radius vector $z(y)$ implies an orthogonal transformation T on the generic point y which leaves h, r, s_{22}, s_{23}, and s_{33} unaltered, and takes the region H into a region H^* such that H and H^* differ on a set of positive measure. T leaves R invariant, hence leaves invariant $\bar{R}$ which uniquely determines the distribution

[4] See footnote 1.

of R. Hence an argument almost the same as that which led us to (3.5) yields the conclusion that the power of H and the power of H^* are equal, identically in B. Proceeding as in Theorem 3, we obtain two essentially different density functions in h, r, θ, s_{22}, s_{23}, s_{33}, whose integrals over the entire space are identical in the elements of B. From these functions we obtain two different density functions in $s_{ij}(i, j = 1, 2, 3)$, with identical moments (obtained by differentiation with respect to the elements of B). The rest of the proof is essentially no different from that of Theorem 3.

THEOREM 8. *In order that the power of H be a function of $\bar{R}$ alone, it is necessary and sufficient that, with the usual exception of a set of measure zero, $V(h, r, s_{22}, s_{23}, s_{33})$ be a function only of h/r* (i.e., *of R*).

The proof of this theorem is essentially the same as the proof of Theorem 4. The place of the transformation Z is taken by one which consists of any linear transformation on the vectors y_2 and y_3, the addition of a constant angle to θ (rotation of $z(y)$), and multiplication of the vector y_1 by a positive scalar c. This transformation leaves $\bar{R}$ invariant. In the proof of sufficiency we use the distribution of R (see, for example, [10, p. 384, equation (15.55)]). The remainder of the proof is essentially the same as that of Theorem 4.

THEOREM 9. *Among all tests H which have the properties described in the conclusions of Theorems 6 and 7, the classical test based on R is the most powerful.*

As a corollary to this theorem we have the following result due to Simaika [3]: Of all tests H whose power is a function of $\bar{R}$ only, the classical test based on R is the most powerful.

Simaika's result also follows easily from Theorem 8 and the density function of R in the same manner that Hsu's result followed from Theorem 4 and the density function of v.

In the course of the proof of Theorem 9, the various symbols W, with or without subscripts, will denote suitable functions of the variables exhibited, and the various symbols k, with or without subscripts, will denote suitable constants.

We have that

$$
\begin{aligned}
P\{H \mid B\} &= \int_H (2\pi)^{(-3n)/2} \mid B \mid^{n/2} \exp\left\{-\tfrac{1}{2}\sum_{j=1}^{n} y_j' B y_j\right\} dy_{11} \cdots dy_{3n} \\
&= \int_H (2\pi\omega^2)^{(-n)/2} \exp\left[-\frac{1}{2\omega^2}\{y_1 - (\beta_2 y_2 + \beta_3 y_3)\}^2\right] \cdot
\end{aligned}
$$

$$
\begin{aligned}
\cdot W_0(s_{22}, s_{23}, s_{33}, \{\sigma_{ij}\})\, dy_{11} \cdots dy_{3n} &= (2\pi\omega^2)^{(-n)/2} \int_H \exp\left\{\frac{1}{\omega^2}(\beta_2 y_2 + \beta_3 y_3)' z\right\} \cdot \\
&\exp\left[-\frac{1}{2\omega^2}\{y_1^2 + \beta_2^2 s_{22} + 2\beta_2\beta_3 s_{23} + \beta_3^2 s_{33}\}\right] \cdot \\
&\cdot W_0(s_{22}, s_{23}, s_{33}, \{\sigma_{ij}\})\, dy_{11} \cdots dy_{3n}.
\end{aligned}
\tag{4.5}
$$

Now $(\beta_2 y_2 + \beta_3 y_3)' z$ is a function only of β_2, β_3, s_{22}, s_{23}, s_{33}, r, and θ. Also

$h^2 + r^2 = s_{11} = y_1^2$. Thus

$$
\begin{aligned}
P\{H \mid B\} &= \int V(h, r, s_{22}, s_{23}, s_{33}) W_1(h, r, s_{22}, s_{23}, s_{33}, \{B\}) \\
&\cdot \exp\left\{\frac{1}{\omega^2}(\beta_2 y_2 + \beta_3 y_3)'z\right\} d\theta\, dh\, dr\, ds_{22}\, ds_{23}\, ds_{33} = \int V(h, r\, s_{22}, s_{23}, s_{33}) \\
&\cdot W_1(h, r, s_{22}, s_{23}, s_{33}, \{B\})(4hr)^{-1} \exp\left\{\frac{1}{\omega^2}(\beta_2 y_2 + \beta_3 y_3)'z\right\} \qquad (4.6)\\
&\cdot d\theta\, dh^2\, dr^2\, ds_{22}\, ds_{23}\, ds_{33} = \int V\left(\sqrt{y_1^2 - r^2}, r, s_{22}, s_{23}, s_{33}\right) \\
&\cdot W_2\left(\sqrt{y_1^2 - r^2}, r, s_{22}, s_{23}, s_{33}, \{B\}\right) \exp\left\{\frac{1}{\omega^2}(\beta_2 y_2 + \beta_3 y_3)'z\right\} \\
&\qquad \cdot d\theta\, dr^2\, dy_1^2\, ds_{22}\, ds_{23}\, ds_{33}.
\end{aligned}
$$

Integrating with respect to θ and designating

$$W_2 \int \exp\left\{\frac{1}{\omega^2}(\beta_2 y_2 + \beta_3 y_3)'z\right\} d\theta$$

by $W(\sqrt{y_1^2 - r^2}, r, s_{22}, s_{23}, s_{33}, \{B\})$ we observe that just as in (2.4), W is monotonically increasing in r (all other variables fixed). Thus we have

$$P\{H \mid B\} = \int VW\, dr^2\, dy_1^2\, ds_{22}\, ds_{23}\, ds_{33}. \tag{4.7}$$

In constructing H only the function V is at our disposal, and this subject to the limitations imposed by the conclusions of Theorems 6 and 7 and the fact that $h^2 + r^2 = y_1^2 = s_{11}$. The function W is not within our control at all. With y_1^2, s_{22}, s_{23}, s_{33} fixed, W is monotonically increasing with r. To maximize the power it is therefore best to distribute the "mass" so that V is as large as possible for large values of r and hence of R. This implies the classical test and proves the theorem.

5. Stringency of the classical tests. Wald [8] calls a test T_1 "most stringent" if the following is true: Let $\{T\}$ be the totality of tests. Let θ be the generic point in the parameter space, and $P\{T \mid \theta\}$ be the power of T at the point θ. Let T_2 be any test other than T_1. Then

$$\sup_{\theta} [\sup_{\{T\}} P\{T \mid \theta\} - P\{T_1 \mid \theta\}] \leq \sup_{\theta} [\sup_{\{T\}} P\{T \mid \theta\} - P\{T_2 \mid \theta\}].$$

Of course, we have omitted to specify the totality $\{T\}$. One can admit all tests whose size $\leq \alpha$, a given constant between 0 and 1, or restrict one's self to tests whose size is exactly α. We shall do the latter.

Under these circumstances we shall prove that the classical test of a linear hypothesis is most stringent. Our proof will occupy but a few lines, and is an easy

consequence of the structure of the classical tests as described in the lemma of section 2. The result itself is a special case of an unpublished theorem due to G. H. Hunt and C. M. Stein, and all priority on this result is theirs.

Return then to the notation of section 2. Let σ be fixed at any arbitrary positive value, and the surface

$$\eta^2 = c_0^2$$

be that one on which

$$\omega_1(\eta) = \sup_{\{T\}} P\{T \mid \eta\} - P\{L_1 \mid \eta\}$$

is a maximum, where L_1 is the classical test of the linear hypothesis. It is clear that this maximum is actually achieved, and that $\omega_1(\eta)$ is a constant on the surface $\eta^2 = c_0^2$. Let L_2 be any other test (of size α), and $\omega_2(\eta)$ be the corresponding function for L_2. We have only to show that on the surface $\eta^2 = c_0^2$ we cannot have everywhere $\omega_2(\eta) < \omega_1(\eta)$, and our proof is complete. If everywhere on the surface $\eta^2 = c_0^2$ we had $\omega_2(\eta) < \omega_1(\eta)$, we would have, also on the same surface, $P\{L_2 \mid \eta\} > P\{L_1 \mid \eta\}$. This would, however, violate Wald's Theorem 2 (section 2) and proves the desired result.

REFERENCES

[1] P. L. Hsu, "Notes on Hotelling's generalized T," *Annals of Math. Stat.*, Vol. 9 (1938) p. 231.

[2] P. L. Hsu, "Analysis of variance from the power function standpoint," *Biometrika*, Vol. 32 (1941), p. 62.

[3] J. B. Simaika, "On an optimum property of two important statistical tests," *Biometrika*, Vol. 32 (1941), p. 70.

[4] A. Wald, "On the power function of the analysis of variance test," *Annals of Math. Stat.*, Vol. 13 (1942), p. 434.

[5] J. A. Shohat and J. D. Tamarkin, *The Problem of Moments*, The American Mathematical Society, New York, 1943.

[6] John Wishart, "The generalized product moment distribution, etc.," *Biometrika*, Vol. 20A (1928), p. 32.

[7] H. Cramér and H. Wold, "Some theorems on distribution functions," *Lond. Math. Soc. Jour.*, Vol. 11 (1936).

[8] A. Wald, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Am. Math. Soc. Trans.*, Vol. 54 (1943), p. 426.

[9] P. C. Tang, "The power function of the analysis of variance etc.," *Stat. Res. Memoirs*, Vol. 2 (1938) (University of London), p. 126.

[10] M. G. Kendall, *The Advanced Theory of Statistics*, Vol. 1, Charles Griffin and Company, London, 1945.

[11] S. Saks, *Theory of the Integral*, (Second Edition), G. E. Stechert and Company, New York, 1937.

[12] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Roy. Soc. London Phil. Trans.*, Ser. A, Vol. 231 (1933), pp. 289–337.

[13] Eberhard Hopf, *Ergodentheorie*, Chelsea, New York, 1948.

APPLICATION OF THE METHOD OF MIXTURES TO QUADRATIC FORMS IN NORMAL VARIATES

By Herbert Robbins and E. J. G. Pitman

Institute of Statistics, University of North Carolina

1. Summary. The method of mixtures, explained in Section 2, is applied to derive the distribution functions of a positive quadratic form in normal variates and of the ratio of two independent forms of this type.

2. The method of mixtures. If

(1) $$F_0(x), \qquad F_1(x),$$

is any sequence of distribution functions, and if

(2) $$c_0, c_1, \cdots$$

is any sequence of constants such that

(3) $$c_j \geq 0 \qquad (j = 0, 1, \cdots), \qquad \Sigma c_j = 1$$

(all summations will be from 0 to ∞ unless otherwise noted), then the function

(4) $$F(x) = \Sigma c_j F_j(x)$$

is called a *mixture* of the sequence (1).

It is sometimes helpful to interpret $F(x)$ in the following manner. Let J, X_0, $X_1, \cdots$ be variates such that J has the distribution $P[J = j] = c_j$ $(j = 0, 1, \cdots)$ and such that X_j has the distribution function $F_j(x)$. Let X be a variate such that the conditional distribution function of X given $J = j$ is $F_j(x)$. Then the distribution function of X is

$$P[X \leq x] = \Sigma P[J = j] \cdot P[X \leq x \mid J = j] = \Sigma c_j F_j(x) = F(x).$$

This interpretation of $F(x)$ will, however, not be involved in the present paper.

The following statements are proved in [1]. If $x = (x_1, \cdots, x_n)$ is a vector variable the function $F(x)$ defined by (4) is a distribution function, and for any Borel set S,

(5) $$\int_S dF(x) = \Sigma c_j \int_S dF_j(x).$$

More generally, if $g(x)$ is any Borel measurable function then

(6) $$\int_{-\infty}^{\infty} g(x)\, dF(x) = \Sigma c_j \int_{-\infty}^{\infty} g(x)\, dF_j(x)$$

whenever the left hand side of (6) exists. In particular, the characteristic function

$\varphi(t)$ corresponding to $F(x)$ is

$$\varphi(t) = \Sigma c_j \varphi_j(t), \tag{7}$$

where $\varphi_j(t)$ is the characteristic function corresponding to $F_j(x)$.

If each $F_j(x)$ has a derivative $f_j(x)$ then $F(x)$ has a derivative $f(x)$ given by

$$f(x) = \Sigma c_j f_j(x), \tag{8}$$

provided that this series converges uniformly in some interval including x. Conversely, if (8) is the relation between the frequency functions and if the series is uniformly convergent in every finite interval, then the relation between the distribution functions is given by (4). In practice we deduce (4) from (8), or, using the uniqueness theorem for characteristic functions, from (7).

As regards computation, we observe that for any integers $0 \leq p_1 \leq p_2$ and for any x it follows from (3) and (4) that

$$\begin{aligned} 0 \leq F(x) - \sum_{p_1}^{p_2} c_j F_j(x) &= \sum_0^{p_1-1} c_j F_j(x) + \sum_{p_2+1}^{\infty} c_j F_j(x) \\ &\leq \sup_{j<p_1} \{F_j(x)\} \cdot \left(\sum_0^{p_1-1} c_j \right) + \sup_{j>p_2} \{F_j(x)\} \cdot \left(1 - \sum_0^{p_1-1} c_j - \sum_{p_1}^{p_2} c_j \right) \leq 1 - \sum_{p_1}^{p_2} c_j. \end{aligned} \tag{9}$$

The existence of these upper bounds (the last a uniform one) for the error term when the series (4) is replaced by a finite sum shows that series expansions of the mixture type (4) are especially well adapted to computational work.

For some purposes it is useful to consider series expansions of the type (4) where the c_j may be of both signs and where the series Σc_j may diverge. Both parts of (3) will, however, be satisfied in the cases considered here.

If U, V are independent variates with respective distribution functions $F(x)$, $G(x)$ we shall denote the distribution function of any Borel measurable function $H(U, V)$ by

$$H(U, V)\ (F(x), G(x)).$$

Now if $F(x)$, $G(x)$ are both mixtures,

$$F(x) = \Sigma c_j F_j(x), \qquad G(x) = \Sigma d_k G_k(x),$$

then by (5),

$$\begin{aligned} P[H(U, V) \leq x] &= \iint_{\{H(u,v) \leq x\}} dF(u)\, dG(v) \\ &= \Sigma\Sigma c_j d_k \iint_{\{H(u,v) \leq x\}} dF_j(u)\, dG_k(v), \end{aligned}$$

so that

$$H(U, V)(\Sigma c_j F_j(x), \Sigma d_k G_k(x)) = \Sigma\Sigma c_j d_k H(u, v)(F_j(x), G_k(x)). \tag{10}$$

As an application of the principles set forth in this section we shall express as series of the mixture type (4) the distribution functions of any positive quadratic form in normal variates and of the ratio of any two independent forms of this type. Special cases of the problem have been dealt with by Tang [2], Hsu [3], and many others, but the method of mixtures permits a unified and simple treatment of the general case.

3. Distribution of a positive quadratic form. We shall denote by $F_n(x)$ the chi-square distribution function with $n > 0$ degrees of freedom,

$$F_n(x) = \frac{1}{2^{\frac{1}{2}n}\cdot\Gamma(\frac{1}{2}n)}\int_0^x u^{\frac{1}{2}n-1}\cdot e^{-\frac{1}{2}u}\cdot du \qquad (x > 0), \tag{11}$$

$$= 0 \qquad (x \le 0)$$

The corresponding characteristic function is

$$\varphi_n(t) = \int_0^\infty e^{ixt}\, dF_n(x) = (1 - 2it)^{-\frac{1}{2}n} = w^{\frac{1}{2}n}, \tag{12}$$

where we have set $w = (1 - 2it)^{-1}$. We shall denote by χ_n^2 any variate with the distribution function (11).

Let a be any constant such that $a > 0$. The characteristic function of the variate $a\cdot\chi_n^2$ is

$$(1 - 2iat)^{-\frac{1}{2}n} = [a(1 - 2it) - (a - 1)]^{-\frac{1}{2}n} = a^{-\frac{1}{2}n}\cdot w^{\frac{1}{2}n}\cdot\left(1 - \left(1 - \frac{1}{a}\right)w\right)^{-\frac{1}{2}n}. \tag{13}$$

By the binomial theorem we have for any $a > 0$,

$$a^{-\frac{1}{2}n}\left[1 - \left(1 - \frac{1}{a}\right)z\right]^{-\frac{1}{2}n} = \Sigma c_j z^j \qquad \left(|z| < \left|1 - \frac{1}{a}\right|^{-1}\right), \tag{14}$$

where

$$c_j = a^{-\frac{1}{2}n}\cdot\frac{\frac{1}{2}n(\frac{1}{2}n + 1)\cdots(\frac{1}{2}n + j - 1)}{j!}\cdot\left(1 - \frac{1}{a}\right)^j \qquad (j = 0, 1, \cdots). \tag{15}$$

For $a \ge 1$ we see from (15) that all the c_j are non-negative. Likewise for $a > \frac{1}{2}$ (and hence *à fortiori* for $a \ge 1$) we have $|1 - 1/a|^{-1} > 1$ so that (14) holds for all $|z| \le 1$; setting $z = 1$ it follows that the sum of all the c_j is equal to 1. Hence for $a \ge 1$,

$$c_j \ge 0 \qquad (j = 0, 1, \cdots), \qquad \Sigma c_j = 1.$$

Since $|w| = |1 - 2it|^{-1} \le 1$ for all real t it follows from (13) and (14) that for $a \ge 1$,

$$(1 - 2iat)^{-\frac{1}{2}n} = \Sigma c_j w^{\frac{1}{2}n+j} = \Sigma c_j(1 - 2it)^{-\frac{1}{2}n-j}$$

$$= \Sigma c_j\, \varphi_{n+2j}(t). \tag{16}$$

Hence for $a \geq 1$ the distribution function $F_n(x/a)$ of the variate $a \cdot \chi_n^2$ is a mixture of χ^2 distribution functions,

$$F_n(x/a) = \Sigma c_j F_{n+2j}(x), \tag{17}$$

where the c_j, determined by the identity (14), are the probabilities of a negative binomial distribution.

It may, in fact be proved by a direct analysis, which we omit here, that (17) holds for any $a > 0$. However, if $a < 1$ then the c_j will be of alternating sign, and if $a \leq \frac{1}{2}$ then the series Σc_j will diverge. This shows incidentally that a relation of the form (4) can hold even though the series Σc_j diverges and hence the corresponding relation (7) does not hold for $t = 0$.

Theorem 1. *Let*

$$X = a(\chi_m^2 + a_1 \chi_{m_1}^2 + \cdots + a_r \chi_{m_r}^2),$$

where the chi-square variates are independent and $a, a_1, \cdots, a_r$ are positive constants such that

$$a_i \geq 1 \qquad (i = 1, \cdots, r).$$

Define constants c_j by the identity[1]

$$\prod_{i=1}^{r} \left\{ a_i^{-\frac{1}{2}m_i} \left[1 - \left(1 - \frac{1}{a_i} \right) z \right]^{-\frac{1}{2}m_i} \right\} = \Sigma c_j z^j \qquad (|z| \leq 1); \tag{18}$$

then obviously

$$c_j \geq 0 \qquad (j = 0, 1, \cdots), \qquad \Sigma c_j = 1.$$

Let

$$M = m + m_1 + \cdots + m_r ;$$

then for every x,

$$P[X \leq x] = \Sigma c_j \cdot F_{M+2j}(x/a). \tag{19}$$

For any integers $0 \leq p_1 \leq p_2$ and every x,

$$\begin{aligned} 0 &\leq P[X \leq x] - \sum_{p_1}^{p_2} c_j F_{M+2j}(x/a) \\ &\leq F_M(x/a) \cdot \left(\sum_{0}^{p_1 - 1} c_j \right) + F_{M+2p_2+2}(x/a) \cdot \left(1 - \sum_{0}^{p_1-1} c_j - \sum_{p_1}^{p_2} c_j \right) \\ &\leq 1 - \sum_{p_1}^{p_2} c_j . \end{aligned} \tag{20}$$

Proof. The characteristic function of X/a is, by (13) and (18),

$$\varphi(t) = w^{\frac{1}{2}M} \cdot \prod_{i=1}^{r} \left\{ a_i^{-\frac{1}{2}m_i} \left[1 - \left(1 - \frac{1}{a_i} \right) w \right]^{-\frac{1}{2}m_i} \right\} = \Sigma c_j w^{\frac{1}{2}M+j} = \Sigma c_j \varphi_{M+2j}(t)$$

[1] If $r = 0$ we regard the left hand side of (18) as having the value 1.

Hence for any y,

$$P[X/a \leq y] = \Sigma c_j F_{M+2j}(y),$$

whence (19) follows on setting $x = ay$. Finally, since $F(x)$ is a decreasing function of n for fixed x, (20) follows from (9).

It should be observed that the coefficients c_j determined by (18) can be written explicitly as the multiple Cauchy products

$$c_j = \sum_{i_1+\cdots+i_r=j} \{c_{1,i_1} \cdots c_{r,i_r}\},$$

where

$$c_{i,j} = a_i^{-\frac{1}{2}m_i} \cdot \frac{\frac{1}{2}m_i(\frac{1}{2}m_i + 1) \cdots (\frac{1}{2}m_i + j - 1)}{j!} \cdot \left(1 - \frac{1}{a_i}\right)^j$$

$$(i = 1, \cdots, r; j = 0, 1, \cdots).$$

The c_j may be computed stepwise by the relations

$$c_j^{(1)} = c_{1,j},$$

$$c_j^{(s)} = \sum_{i=0}^{j} \{c_{j-i}^{(s-1)} \cdot c_{s,i}\} \qquad (s = 2, \cdots, r),$$

$$c_j^{(r)} = c_j.$$

4. Distribution of a ratio. The ratio χ_m^2/χ_n^2 of two independent chi-square variates has the distribution function

$$F_{m,n}(x) = \frac{\Gamma(\frac{1}{2}(m+n))}{\Gamma(\frac{1}{2}m)\Gamma(\frac{1}{2}n)} \int_0^x u^{\frac{1}{2}m-1} (1+u)^{-\frac{1}{2}(m+n)} \, du \qquad (x \geq 0), \tag{21}$$

$$= 0 \qquad (x < 0).$$

In computational work we can use the tables of the Beta distribution function

$$I_x(r, s) = \frac{\Gamma(r+s)}{\Gamma(r)\cdot\Gamma(s)} \int_0^x u^{r-1} \cdot (1-u)^{s-1} \cdot du \qquad (0 < x < 1),$$

$$= 0 \ (x \leq 0), \qquad 1 \ (x \geq 1),$$

together with the identity

$$F_{m,n}(x) = I_{x/(1+x)}(\tfrac{1}{2}m, \tfrac{1}{2}n).$$

THEOREM 2. *Let*

$$X = \frac{a\cdot(\chi_m^2 + a_1\chi_{m_1}^2 + \cdots + a_r\chi_{m_r}^2)}{\chi_n^2 + b_1\chi_{n_1}^2 + \cdots + b_s\chi_{n_s}^2}, \tag{22}$$

where the χ^2 variates are independent and $a, a_1, \cdots, a_r, b_1, \cdots, b_s$ are positive

constants such that

$$a_i \geq 1, \qquad b_j \geq 1$$

$$(i = 1, \cdots, r; j = 1, \cdots, s).$$

Define constants c_j, d_k *by the identities*

$$\prod_{i=1}^{r} \left\{ a_i^{-\frac{1}{2}m_i} \cdot \left[1 - \left(1 - \frac{1}{a_i} \right) z \right]^{-\frac{1}{2}m_i} \right\} = \Sigma c_j z^j, \qquad (|z| \leq 1)$$

$$\prod_{i=1}^{s} \left\{ b_i^{-\frac{1}{2}n_i} \cdot \left[1 - \left(1 - \frac{1}{b_i} \right) z \right]^{-\frac{1}{2}n_i} \right\} = \Sigma\, d_k z^k;$$

then

$$c_j \geq 0, \quad \Sigma c_j = 1, \qquad d_k \geq 0, \quad \Sigma d_k = 1.$$

Let

$$M = m + m_1 + \cdots + m_r, \qquad N = n + n_1 + \cdots + n_s;$$

then for every x,

$$P[X \leq x] = \Sigma\Sigma c_j\, d_k \cdot F_{M+2j, N+2k}(x/a),$$

and for any integers $0 \leq p_1 \leq p_2$, $0 \leq q_1 \leq q_2$ *and every* x,

$$0 \leq P[X \leq x] - \sum_{p_1}^{p_2} \sum_{q_1}^{q_2} c_j\, d_k \cdot F_{M+2j, N+2k}(x/a)$$

$$\leq \left(1 - \sum_{p_1}^{p_2} c_j \right) \cdot \left(1 - \sum_{q_1}^{q_2} d_k \right).$$

PROOF. Let U, V denote respectively numerator and denominator of (22). From Theorem 1,

$$P[U \leq x] = \Sigma c_j\, F_{M+2j}(x/a),$$

$$P[V \leq x] = \Sigma\, d_k\, F_{N+2k}(x).$$

Hence by (10), for every x,

$$P[X \leq x] = P[U/V \leq x] = \Sigma\Sigma c_j\, d_k \cdot F_{M+2j, N+2k}(x/a).$$

The rest of the theorem is obvious.

COROLLARY. *Let*

$$X = \frac{\chi_M^2}{a\chi_r^2 + b\chi_s^2},$$

where the χ^2 *variates are independent and*

$$0 < a \leq b.$$

Define

$$\alpha = a/b, \qquad N = r + s,$$

$$c_j = \alpha^{\frac{1}{2}s} \cdot \frac{\frac{1}{2}s(\frac{1}{2}s+1)\cdots(\frac{1}{2}s+j-1)}{j!} \cdot (1-\alpha)^j \qquad (j = 0, 1, \cdots);$$

then

$$c_j \geq 0 \qquad (j = 0, 1, \cdots), \qquad \Sigma c_j = 1,$$

and for every x,

$$P[X \leq x] = \Sigma c_j F_{M,N+2j}(ax).$$

For any integers $0 \leq p_1 \leq p_2$ and every x,

$$0 \leq p[X > x] - \sum_{p_1}^{p_2} c_j[1 - F_{M,N+2j}(ax)]$$

$$\leq [1 - F_{M,N}(ax)] \cdot \left(\sum_0^{p_1-1} c_j\right) + [1 - F_{M,N+2p_2+2}(ax)] \tag{23}$$

$$\cdot\left(1 - \sum_0^{p_1-1} c_j - \sum_{p_1}^{p_2} c_j\right) \leq 1 - \sum_{p_1}^{p_2} c_j.$$

PROOF. Except for (23) this is a special case of Theorem 2. To prove (23) we observe that

$$P[X > x] = 1 - P[X \leq x] = \Sigma c_j[1 - F_{M,N+2j}(ax)],$$

and since for fixed m and x, $F_{m,n}(x)$ is an increasing function of n, (23) follows in the same way as (9).

5. The non-central case. Let Y be normal (0, 1) and let $X = (Y + d)^2$, where d is any constant. The frequency function of X is, for $x > 0$,

$$f(x) = (2\pi x)^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(d^2+x)} \cdot (e^{dx^{\frac{1}{2}}} + e^{-dx^{\frac{1}{2}}})/2.$$

By expanding the last factor into a power series it is easily seen that

$$f(x) = \Sigma p_j \cdot f_{1+2j}(x), \tag{24}$$

where $f_n(x) = F'_n(x)$ is the chi-square frequency function with n degrees of freedom and where

$$p_j = e^{-\frac{1}{2}d^2} \cdot (\tfrac{1}{2}d^2)^j/j! \qquad (j = 0, 1, \cdots).$$

Since the identity

$$e^{-\frac{1}{2}d^2(1-z)} = \Sigma p_j z^j \qquad \text{(all } z\text{)} \tag{25}$$

holds, it follows that

$$p_j \geq 0 \qquad (j = 0, 1, \cdots), \qquad \Sigma p_j = 1.$$

The series (24) is uniformly convergent in every finite interval, so that we can write the distribution function $F(x)$ and characteristic function $\varphi(t)$ of X in the forms

$$F(x) = \Sigma p_j \cdot F_{1+2j}(x),$$
$$\varphi(t) = \Sigma p_j \cdot \varphi_{1+2j}(t) = w^{\frac{1}{2}} \cdot e^{-\frac{1}{2}d^2(1-w)},$$

where again we have set $w = (1 - 2it)^{-1}$.

Now let $Y_1, \cdots, Y_n$ be independent and normal (0, 1) variates and let

$$X = (Y_1 + d_1)^2 + \cdots + (Y_n + d_n)^2, \tag{26}$$

where the d_i are constants such that

$$d_1^2 + \cdots + d_n^2 = d^2.$$

The characteristic function of X is then

$$\varphi(t) = w^{\frac{1}{2}n} \cdot e^{-\frac{1}{2}d^2(1-w)} = \Sigma p_j\, w^{\frac{1}{2}n+j} = \Sigma p_j\, \varphi_{n+2j}(t),$$

and hence the distribution function $F(x)$ of X is again a mixture of χ^2 distribution functions,

$$F(x) = \Sigma p_j \cdot F_{n+2j}(x), \tag{27}$$

where the p_j, determined by the identity (25), are the probabilities of a Poisson distribution with parameter $\lambda = \frac{1}{2}d^2$. We shall denote the non-central chi-square variate (26) by $\chi'^2_{n,d}$.

We can now generalize Theorems 1 and 2 in a straightforward manner to cover non-central chi-square variates. We shall state only the generalization of the Corollary of Theorem 2 to the case in which the numerator is non-central.

THEOREM 3. *Let*

$$X = \frac{\chi'^2_{M,d}}{a\chi_r^2 + b\chi_s^2},$$

where the χ^2 *variates are independent and*

$$0 < a \leq b.$$

Define

$$\lambda = \tfrac{1}{2}d^2, \qquad \alpha = a/b, \qquad N = r + s$$

$$p_j = e^{-\lambda} \cdot \lambda^j / j! \qquad (j = 0, 1, \cdots),$$

$$c_k = \alpha^{\frac{1}{2}s} \cdot \frac{\frac{1}{2}s(\frac{1}{2}s + 1) \cdots (\frac{1}{2}s + k - 1)}{k!} \cdot (1 - \alpha)^k \qquad (k = 0, 1, \cdots);$$

then

$$p_j \geq 0, \quad \Sigma p_j = 1, \qquad c_k \geq 0, \quad \Sigma c_k = 1,$$

and for every x,

$$P[X \leq x] = \Sigma\Sigma p_j c_k F_{M+2j,N+2k}(ax).$$

For any integers $0 \leq g_1 \leq g_2$, $0 \leq h_1 \leq h_2$,

$$0 \leq P[X \leq x] - \sum_{g_1}^{g_2} \sum_{h_1}^{h_2} p_j c_k \cdot F_{M+2j,N+2k}(ax) \leq \left(1 - \sum_{g_1}^{g_2} p_j\right) \cdot \left(1 - \sum_{h_1}^{h_2} c_k\right).$$

REFERENCES

[1] HERBERT ROBBINS, "Mixture of distributions," *Annals of Math. Statistics*, Vol. 19 (1948), p. 360.

[2] P. C. TANG, "The power function of the analysis of variance tests with tables and illustrations of their use," *Stat. Res. Mem.*, Vol. 2 (1938), p. 126.

[3] P. L. HSU, "Contributions to the theory of 'Student's' t-test as applied to the problem of two samples," *ibid.*, p. 1.

THE JOINT DISTRIBUTION OF SERIAL CORRELATION COEFFICIENTS

By M. H. Quenouille

Rothamsted Experimental Station

1. Summary. An expression for the joint distribution of serial correlation coefficients, circularly defined, has been derived. It has been shown that this distribution possesses properties similar to those already encountered in the distribution of a single serial correlation coefficient, i.e. it is definedby different function forms for various subregions. The distribution thus found is of little use for computational purposes. Consequently, approximate forms have been investigated and the suitability of the ordinary partial correlation coefficient for large-sample testing has been inferred.

2. Introduction. Anderson [1] has derived the distribution of the serial correlation coefficient

$$r_l = \frac{\sum_{i=1}^{n} \epsilon_i \epsilon_{i+l} - \left(\sum_{i=1}^{n} \epsilon_i\right)^2 \Big/ n}{\sum_{i=1}^{n} \epsilon_i^2 - \left(\sum_{i=1}^{n} \epsilon_i\right)^2 \Big/ n},$$

where the ϵ_i are normally and independently distributed with mean μ and variance σ^2 and where a circular definition is employed, so that ϵ_{n+i} is defined to be equal to ϵ_i. However, in making a test of any series, we shall usually be faced with a set of serial correlation coefficients, so that we shall require a joint distribution function of $r_1, r_2, \cdots, r_m$ say. This distribution function is derived below by an extension of the method used by Koopmans [2].

It should be noted that Bartlett [3] has shown that for large samples the variances and covariances of the r_l are independent of the distribution of ϵ_i under fairly wide conditions. This means that the joint distribution function obtained for normal ϵ_i will often give a good approximation for non-normal ϵ_i and can be used as the basis for any test of the correlogram.

3. Conditions on the r_l. It is easily seen that the r_l cannot take all values from $+1$ to -1 independently. For example, r_2 cannot take a value near -1 if r_l takes a value near $+1$. As a result, there will be certain necessary conditions that the r_l will have to fulfil. It is not difficult to find these conditions, since, if $y_i (i = 1, 2, \cdots, n)$ are any set of variables, then

$$\sum_{j=1}^{n} (\epsilon_{i+j} y_i)^2 = \left(\sum_{i=1}^{n} \epsilon_i^2\right) r_j y_l y_{l+j}, \tag{1}$$

where ϵ_i may or may not be corrected for the mean and the double-suffix summation convention is employed.

Thus, provided $0 < m < n/2$, we will have

$$
\text{(2)} \qquad R_m = \begin{vmatrix} 1 & r_1 & r_2 & \cdots & r_m \\ r_1 & 1 & r_1 & \cdots & r_{m-1} \\ r_2 & r_1 & 1 & \cdots & r_{m-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_m & r_{m-1} & r_{m-2} & \cdots & 1 \end{vmatrix} \geq 0
$$

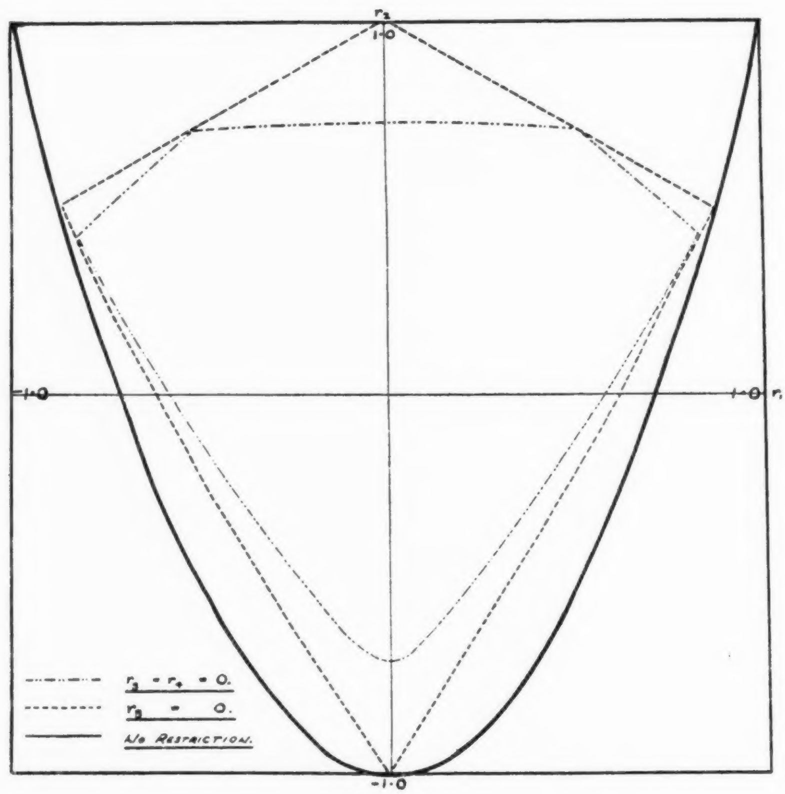

Fig. 1

as a necessary condition that the right-hand side of (1) be positive definite and this expression will impose necessary conditions upon the joint distribution of the r_l.

Fig. 1 gives the limits of possible values of r_1 and r_2 subject to (a) no restriction, (b) $r_3 = 0$, (c) $r_3 = r_4 = 0$.

4. Complex Integration in m Variables. Before finding the joint distribution function of the r_i some introductory remarks on complex integration involving m variables will be necessary.

We can evaluate an integral such as

$$\int_{-\infty}^{\infty} \cdots \int \frac{f(z_1, z_2 \cdots z_m)}{\prod_{j=1}^{m} (z_j - a_j)} dz_1 \cdots dz_m$$

where $\mathscr{I}(a_i) = 0$ and $f(z_1, z_2, \cdots, z_m)$ is regular in the region $\mathscr{I}(z_i) \geq 0$, by successive Cauchy integrations, so the integral has a value $(2\pi i)^m f(a_1, \cdots, a_m)$. In the same manner as for Cauchy integration, it will be possible to distort the contours over which we integrate so that we can evaluate

$$\int_S \cdots \int \frac{f(z_1 \cdots z_m)}{\prod_{j=1}^{m} (z_j - a_j)} dz_1 \cdots dz_m,$$

provided that $f(z_1, \cdots, z_m)$ is regular in the region defined by S, and $(a_1, \cdots, a_m)$ is enclosed in this region.

More generally, if we have an integral of the form

$$\int_S \cdots \int \frac{f(z_1 \cdots z_m)}{\prod_{j=1}^{m} (a_{ij} z_i - b_j)} dz_1 \cdots dz_m,$$

and we make the transformations $w_j = a_{ij} z_i$ and $b_j = a_{ij} c_i$, i.e. $\boldsymbol{W} = \boldsymbol{AZ}$, $\boldsymbol{C} = \boldsymbol{A}^{-1}\boldsymbol{B}$, it is possible, in the above manner, to evaluate the integral as

$$\pm \frac{(2\pi i)^m}{|\boldsymbol{A}|} f(c_1 \cdots c_m). \tag{3}$$

Suppose we now consider the integral

$$\int_S \cdots \int \frac{f(z_1 \cdots z_m)}{\prod_{j=1}^{n} (a_{ij} z_i - b_j)} dz_1 \cdots dz_m,$$

where $n \geq m$. We may select a set, g_k, of m equations $a_{ij} z_i = b_j$, and let $\boldsymbol{A}_k = [a_{ij}]$, $\boldsymbol{B}_k = [b_j]$, $\boldsymbol{C}_k = \boldsymbol{A}_k^{-1}\boldsymbol{B}_k = [c_{ik}]$. Then, we may carry out the integration as previously, in this case, summing a series of terms for various combinations of m equations out of the possible n. The value of the integral may then be written

$$(2\pi i)^m \sum_{g_k} \pm \frac{f(c_{1k}, \cdots c_{nk})}{|\boldsymbol{A}_k| \prod_{l \neq g_k} (a_{lj} c_{lk} - b_j)}, \tag{4}$$

where the summation occurs over the points $(c_{ik}, c_{2k}, \cdots, c_{nk})$ lying in the region defined by S, and the product term excludes the set of equations g_k. The ambiguity of sign in (3) and (4) arises from the Jacobian $|\boldsymbol{A}_k|^{-1}$, and the sign must be chosen which makes the transformation of $dz_1, \cdots, dz_m$ yield a positive

element. It must be noted that it is possible to obtain several expansions of the form (4) according to the convention that is employed in defining "enclosure" for each of the variables.

5. Integral form for the joint distribution function. We can, without loss of generality, assume $\sigma^2 = 1$. Suppose that

$$p = \sum_{i=1}^{n} \epsilon_i^2 - \left(\sum_{i=1}^{n} \epsilon_i\right)^2 \Big/ n, \qquad q_l = \sum_{i=1}^{n} \epsilon_i \epsilon_{i+l} - \left(\sum_{i=1}^{n} \epsilon_i\right)^2 \Big/ n,$$

where $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ are independent, so that $r_l = q_l/p$. Then by a consideration of n dimensional space, we can see that p is distributed independently of $r_1, \cdots r_m$ so that their joint distribution can be written $g(p)h(r_1, \cdots, r_m)dp\, dr_1, \cdots, dr_m$. The joint distribution of p and $q_1, \cdots, q_m$ can thus be written

$$f(pq_1 \cdots q_m)\, dp\, dq_1 \cdots dq_m = \frac{g(p)}{p^m}\, h\left(\frac{q_1}{p}, \cdots \frac{q_m}{p}\right) dp\, dq_1 \cdots dq_m, \tag{5}$$

where it is not difficult to see that

$$g(p) = \frac{p^{\frac{1}{2}(n-3)} e^{-\frac{1}{2}p}}{2^{\frac{1}{2}(n-1)} \Gamma\left(\dfrac{n-1}{2}\right)}. \tag{6}$$

We can now find the joint distribution of p and $q_1, \cdots, q_m$ by inverting the characteristic function of these variables. This is given by

$$\frac{1}{(2\pi)^{\frac{1}{2}n}} \int_{-\infty}^{\infty} \cdots \int \exp\left[-\frac{\Sigma\epsilon_j^2}{2} + i(\eta p + \theta_j q_j)\right] d\epsilon_1 \cdots d\epsilon_n,$$

$$= \frac{1}{(2\pi)^{\frac{1}{2}n}} \int_{-\infty}^{\infty} \cdots \int \exp\left[-\frac{\epsilon' \Delta \epsilon}{2}\right] d\epsilon_1 \cdots d\epsilon_n,$$

$$= 1/|\Delta|^{\frac{1}{2}},$$

where

$$\epsilon' = [\epsilon_1, \epsilon_2, \cdots, \epsilon_n]$$

and

$$|\Delta| = \prod_{l=1}^{n-1} (1 - 2i\eta - 2i\theta_j \kappa_{jl}), \qquad \kappa_{jl} = \cos \frac{2\pi_j\, l}{n},$$

$$= (1 - 2i\eta)^{n-1} \prod_{l=1}^{n-1} (1 - \kappa_j \kappa_{jl}), \qquad \kappa_j = \frac{2i\theta_j}{1 - 2i\eta},$$

so that the joint distribution of p and $q_1, \cdots, q_m$ is

$$f(p, q_1 \cdots q_m) = \frac{1}{(2\pi)^{m+1}} \int_{-\infty}^{\infty} \cdots \int \frac{1}{|\Delta|^{\frac{1}{2}}} \exp\{-i(\eta p + \theta_j q_j)\}\, d\eta\, d\theta_1 \cdots d\theta_m$$

$$= \frac{1}{(2\pi)^{m+1}} \int_{-\infty}^{\infty} e^{-i\eta p}(1 - 2i\eta)^{-\frac{1}{2}(n-2m-1)} \int_{S_\eta} \frac{1}{|\Delta|^{\frac{1}{2}}} \exp\left\{-\frac{(1 - 2i\eta)\kappa_j q_j}{2}\right\} \frac{d\kappa_1 \cdots d\kappa_m}{(2i)^m}\, d\eta, \tag{7}$$

where S_η is the region bounded by $\kappa_j = \pm \dfrac{2i\infty}{1 - 2i\eta}$. Now S_η can be replaced by region S enclosing the same set of singularities on the real hyperplane, and S can be chosen independent of η. Thus it will be possible to reverse the order of integration in (7) provided that $\int_{-\infty}^{\infty} | 1 - 2i\eta |^{-\frac{1}{2}(n-2m-1)} d\eta$ converges, i.e. provided $n > 2m + 3$. Then since

$$\begin{aligned} \frac{1}{2\pi} \int_{-\infty}^{\infty} (1 - 2i\eta)^{-\frac{1}{2}(n-2m-1)} \exp \{-i\eta(p - \kappa_j q_j)\}\, d\eta \\ &= \frac{(p - \kappa_j q_j)^{\frac{1}{2}(n-2m-3)}}{2^{\frac{1}{2}(n-2m-1)} \Gamma\left(\dfrac{n - 2m - 1}{2}\right)} \exp \{-\tfrac{1}{2}(p - \kappa_j q_j)\} \quad \text{for} \quad p \geq \kappa_j q_j, \\ &= 0 \quad \text{for} \quad p \leq \kappa_j q_j, \end{aligned}$$

we get

$$f(p, q_1 \cdots q_m) = \frac{e^{-\frac{1}{2}p}}{2^{\frac{1}{2}(n-1)} (2\pi i)^m \Gamma\left(\dfrac{n - 2m - 1}{2}\right)} \cdot \int_S \cdots \int \frac{(p - \kappa_j q_j)^{\frac{1}{2}(n-2m-3)}}{\left[\prod_{l=1}^{n-1} (1 - \kappa_j \kappa_{jl})\right]^{\frac{1}{2}}} d\kappa_1 \cdots d\kappa_m, \tag{8}$$

where S encloses the same singularities as S_η all of which lie in the region $p \geq \kappa_j q_j$. If we now use (5) and (6) we get

$$h(r_1 \cdots r_m) = \frac{\Gamma\left(\dfrac{n - 1}{2}\right)}{\Gamma\left(\dfrac{n - 2m - 1}{2}\right)} \cdot \frac{1}{(2\pi i)^m} \int_S \cdots \int \frac{(1 - \kappa_j r_j)^{\frac{1}{2}(n-2m-3)}}{\left[\prod_{l=1}^{n-1} (1 - \kappa_j \kappa_{jl})\right]^{\frac{1}{2}}} d\kappa_1 \cdots d\kappa_m. \tag{9}$$

In a similar manner, it is possible to derive for $n \geq 2m + 3$ the joint distribution of serial correlation coefficients, $\bar{r}_1, \cdots, \bar{r}_m$, uncorrected for the mean, in the form

$$h(\bar{r}_1 \cdots \bar{r}_m) = \frac{\Gamma\left(\dfrac{n}{2}\right)}{\Gamma\left(\dfrac{n - 2m}{2}\right)} \frac{1}{(2\pi i)^m} \int_S \cdots \int \frac{(1 - \kappa_j \bar{r}_j)^{\frac{1}{2}(n-2m-2)}}{\left[\prod_{l=1}^{n} (1 - \kappa_j \kappa_{jl})\right]^{\frac{1}{2}}} d\kappa_1 \cdots d\kappa_m. \tag{10}$$

6. Extension for variables in an autoregressive scheme. Madow [4] has shown how to extend the distribution of the serial correlation coefficient for uncorrelated variables to the case when the variables x_i are connected by a linear Markoff scheme, $x_i = \rho x_{i-1} + \epsilon_i$ with a normal distribution of the error ϵ_i. It is worth

noting that the method used by Madow can be applied to derive the joint distribution of serial correlations of variables x_i, which are connected by a linear autoregressive scheme of order m, or less,

$$a_0 x_i + a_1 x_{i-1} + \cdots + a_m x_{i-m} = \epsilon_i,$$

where $\epsilon_1, \cdots, \epsilon_n$ are normally and independently distributed, and $\epsilon_{n+i} = \epsilon_i$.[1] Under these conditions, the expression (9) will be modified by a factor

$$\left(\frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} \epsilon_i^2}\right)^{\frac{1}{2}(n-1)} = \frac{1}{(A + 2B_j r_j)^{\frac{1}{2}(n-1)}}, \tag{11}$$

where

$$A = \sum_{k=0}^{m} a_k^2,$$

$$B_j = \sum_{k=0}^{m-j} a_k a_{k+j},$$

while (10) will be modified by a similar factor with n replacing $n - 1$.

7. Reduction of the distribution function integral. Using the method described in section 4, it is now possible to reduce the integral given in (9), if we observe that $\kappa_{jl} = \kappa_{jn-l}$ and assume n odd. We then have

$$\begin{aligned} h(r_1 \cdots r_m) &= \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2m-1}{2}\right)} \frac{1}{(2\pi i)^m} \int_S \cdots \int \frac{(1 - \kappa_j r_j)^{\frac{1}{2}(n-2m-3)}}{\prod_{l=1}^{\frac{1}{2}(n-1)} (1 - \kappa_j \kappa_{jl})} d\kappa_1 \cdots d\kappa_m \\ &= \frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-2m-1}{2}\right)} \sum_{g_k} \frac{\begin{vmatrix} 1 & I \\ r & K_k \end{vmatrix}^{\frac{1}{2}(n-2m-3)}}{\prod_{l \neq g_k} \begin{vmatrix} 1 & I \\ \kappa_{jl} & K_k \end{vmatrix}}, \end{aligned} \tag{12}$$

where $\boldsymbol{I} = (1, 1, \cdots, 1)$, $\boldsymbol{r}' \doteq (r_1, r_2, \cdots, r_m)$, $\underline{\kappa}_{jl} = (\kappa_{jl}, \cdots, \kappa_{ml})$ and K_k is the matrix formed from a set g_k of the m matrices $\underline{\kappa}_{jl}$ arranged in order. The factors in the summation can most easily be determined if we put $\begin{vmatrix} 1 & I \\ \boldsymbol{r} & \boldsymbol{K}_k \end{vmatrix} \propto A(r_1, \cdots, r_{m-1}) - r_m$ and sum over the region for which $r_m \leq A(r_1, \cdots, r_{m-1})$. To demonstrate the manner in which formula (12) works, we shall consider $m = 2$. From formula (2) we can see that a limit to the possible values that r_2 can take is given by $r_2 = 2r_1^2 - 1$ i.e. by the curve $(\cos\theta, \cos 2\theta)$

[1] This is a sufficient condition for $x_{n+i} = x_i$.

in the (r_1, r_2) plane. It is not difficult to see that there are ${}^nC_2$ possible terms in (12) and that each of these terms is proportional to the $\frac{1}{2}(n - 2m - 3)$th power of the distance from a line in the (r_1, r_2) planes. These lines are the joins of the points $(\cos 2\pi i/n, \cos 4\pi i/n)$, $i = 1, \cdots, \frac{1}{2}(n - 1)$ and the joins of such points on the curve $(\cos \theta, \cos 2\theta)$ give the outer limits of the possible values of r_1 and r_2. It can also be seen that these points correspond to the equations $\kappa_j \kappa_{jl} = 1$ (each of these equations determines a plane in 4-dimensional complex space), while the joins of these points correspond to the singularities defined by and terms arising from pairs of these equations. Furthermore, since the sum of residues in any plane is zero, the sum of contributions, taken with appropriate signs, arising from lines through any of these points is zero, i.e. the sum of all possible terms involving any particular κ_{jl} will disappear. This leads to several possible expansions for $h(r_1, \cdots, r_m)$.

If we consider the particular case $n = 9$, then each term in the expansion (12) is proportional to the distance from one of the lines joining $(\cos 2\pi i/9 \cos 4\pi i/9)$, $i = 1, 2, 3, 4$. These lines may be denoted by l_{ij}. Then the contribution from l_{ij} is given by

$$3 \frac{\kappa_{1i}\kappa_{1j} - (\kappa_{1i} + \kappa_{1j})r_1 + \frac{1}{2}(r_2 + 1)}{(\kappa_{1j} - \kappa_{1i})(\kappa_{1i} - \kappa_{1k})(\kappa_{1i} - \kappa_{1l})(\kappa_{1j} - \kappa_{1k})(\kappa_{1j} - \kappa_{1l})},$$

where $j > i$ and $\kappa_{1\alpha} = \cos \dfrac{2\pi\alpha}{q}$.

The values of this expression are:

$$\begin{aligned}
&l_{12}, && -1.979 + 2.938\, r_1 - 1.563\, r_2,\\
&l_{13}, && 0.926 - 2.106\, r_1 + 3.959\, r_2,\\
&l_{14}, && 1.053 - 0.832\, r_1 - 2.396\, r_2,\\
&l_{23}, && -5.012 - 3.959\, r_1 - 6.065\, r_2,\\
&l_{24}, && 3.033 + 6.897\, r_1 + 4.502\, r_2,\\
&l_{34}, && -4.086 - 6.065\, r_1 - 2.106\, r_2,
\end{aligned}$$

where, for example, the contribution from l_{12} acts in the region for which $1.563\, r_2 \leq -1.979 + 2.938\, r_1$. Fig. 2 demonstrates the configuration for this case. It is seen that the frequency surface is a tetrahedron. As particular examples of the identities mentioned above we have

$$\begin{aligned}
l_{12} + l_{13} + l_{14} &= 0,\\
-l_{12} + l_{23} + l_{24} &= 0,\\
-l_{13} - l_{23} + l_{34} &= 0.
\end{aligned}$$

For a general value of m, we shall find that the hyperplanes joining sets of m points $(\cos 2\pi i/n, \cos 4\pi i/n, \cdots, \cos 2\pi mi/n)$ will be singularities on the

frequency hypersurface. The hyperplanes passing through sets of m successive points will give the limits of possible values of $r_1, \cdots, r_m$. Furthermore, the sum of contributions (with appropriate signs) to the frequency function from the set of $\frac{1}{2}(n - 2m + 1)$ hyperplanes passing through any point will be zero.

8. Integral approximation for the distribution function. The expression (12) is, of course, difficult to use in practice and we require an approximation similar to that of Koopmans. For this we make use of the integral expression (10)

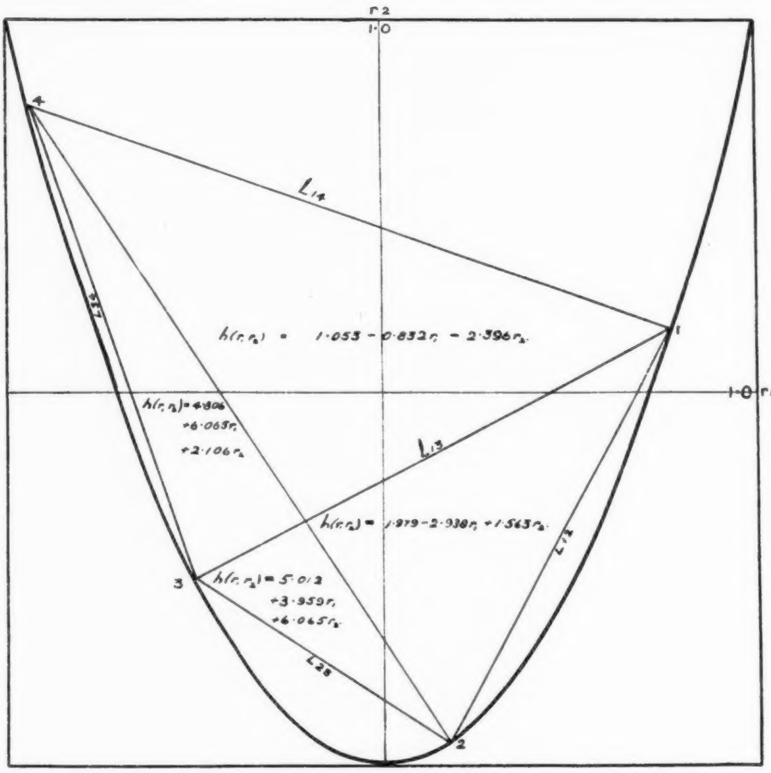

FIG. 2

for the joint distribution function of $\bar{r}_1, \cdots, \bar{r}_m$ and approximate to the factor $\left[\prod_{l=1}^{n} (1 - \kappa_j \kappa_{jl})\right]^{-\frac{1}{2}}$. This can be done without undue difficulty, but the resulting multiple integral does not appear to be capable of easy reduction. This is hardly surprising, since from the nature of the distribution of the r_i we should expect this approximation to involve R_m raised to a suitable power, and this conjecture is strengthened by the following considerations:

a) The distribution of $\bar{r}_1$ may be obtained by considering the two sets of observations $x_1, x_2, \cdots, x_{n-1}, x_n$ and $x_2, x_3, \cdots, x_n, x_1$ as unrelated, and using

the distribution of the ordinary correlation coefficient corresponding to $n + 3$ pairs of observations. (Dixon [6] Quenouille [7]). In the same manner, the m sets of observations $x_1, x_2, \cdots, x_{n-1}, x_n; x_2, x_3, \cdots, x_n; \cdots x_m, x_{m+1}, \cdots, x_{m-2}, x_{m-1}$, might be considered as unrelated and the joint distribution of their correlations, given by Garding (5), will involve R_m raised to a suitable power.
b) The outer limits for the joint distribution of $r_1, r_2, \cdots, r_m$ or $\bar{r}_1, \bar{r}_2, \cdots, \bar{r}_m$ for large n, will be provided by the equations $R_p = 0$, $(p = 1, \cdots, m)$. An investigation of the properties of the functions, $R_1, R_2, \cdots, R_m$ might therefore be expected to throw light upon the joint distribution of $r_1, r_2, \cdots, r_m$ or $\bar{r}_1, \bar{r}_2, \cdots, \bar{r}_m$.
c) R_p is a quadratic in r_p and may be put equal to $R_{p-2}(r'_p - r_p)(r_p - r''_p)$, where r'_p and r''_p are functions of $r_1, r_2, \cdots, r_{p-1}$ giving the limits of the values that r_p can take for any particular values of $r_1, \cdots, r_{p-1}$. Let $Q_p = R_p/R_{p-1}$, then Q_p is likewise a quadratic in r_p, taking all values between r'_p and r''_p and

$$\int_{r''_p}^{r'_p} Q_p^s \, dr_p = \frac{R_{p-2}^s}{R_{p-1}} \int_{r''_p}^{r'_p} (r'_p - r_p)^s (r_p - r''_p)^s \, dr_p$$
$$= \frac{B(s + 1, \frac{1}{2})}{Q_{p-1}^s} \cdot \left(\frac{r'_p - r''_p}{2}\right)^{2s+1}.$$

But, by expanding R_p as a bordered determinant, it is not difficult to show that $r'_p - r''_p = 2Q_{p-1}$, so that

$$\int_{r''_p}^{r'_p} Q_p^s \, dr_p = \frac{\Gamma(s + 1)}{\Gamma(s + \frac{3}{2})} \cdot \pi^{\frac{1}{2}} \cdot Q_{p-1}^{s+1}.$$

In particular, if

$$f(r_1 \cdots r_m) = \frac{\Gamma(\frac{1}{2}n + 1) \cdots \Gamma(\frac{1}{2}n - m + 2)}{\Gamma(\frac{1}{2}n + \frac{1}{2}) \cdots \Gamma(\frac{1}{2}n - m + \frac{3}{2})} \cdot \frac{1}{\pi^{m/2}} Q_m^{\frac{1}{2}(n-2m+1)}, \tag{13}$$

and if we integrate with respect to $r_m, r_{m-1}, \cdots r_2$ in turn, we get

$$\int_{r''_2}^{r} \cdots \int_{r''_m}^{r'_m} f(r_1 \cdots r_m) \, dr_m \cdots dr_2 = \frac{\Gamma(\frac{1}{2}n + 1)}{\Gamma(\frac{1}{2}n + \frac{1}{2})\pi^{\frac{1}{2}}} (1 - r^2)^{\frac{1}{2}(n-1)},$$

which is the approximate distribution of the first serial correlation coefficient, uncorrected for the mean, as given by Dixon [6].

The importance of this lies in the fact that the integral corresponding to that of Koopman's for the joint distribution is

$$\frac{\Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2}n - m)} \cdot \frac{2^{\frac{1}{2}n}}{\Pi^m} \int_S \cdots \int \frac{\begin{vmatrix} 1 & \underline{I} \\ r & \mathbf{X} \end{vmatrix}^{\frac{1}{2}n-m-1}}{|\mathbf{Y}|^{\frac{1}{2}n}} \prod_{i=1}^{m} \left[\sin \tfrac{1}{2}nx_i \begin{vmatrix} & 0 & \\ \dfrac{d}{dx_i} & \kappa(x_i) & \begin{matrix} I \\ X \end{matrix} \end{vmatrix} \right] dx_1 \cdots dx_m$$

where $\boldsymbol{r}' = [r_1, \cdots, r_m]$,

$$\mathbf{X} = \begin{bmatrix} \cos x_1 & \cos x_2 & \cdots & \cos x_m \\ \cos 2x_1 & \cos 2x_2 & \cdots & \cos 2x_m \\ \cdots & \cdots & \cdots & \cdots \\ \cos mx_1 & \cos mx_2 & \cdots & \cos mx_m \end{bmatrix},$$

$$\mathbf{I} = [1, 1, \cdots 1],$$

$$\mathbf{Y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \cos x_1 & \cos x_2 & \cdots & \cos x_m \\ \cdots & \cdots & \cdots & \cdots \\ \cos (m-1)x_1 & \cos (m-1)x_2 & \cdots & \cos (m-1)x_m \end{bmatrix},$$

$$\boldsymbol{\kappa}'(\theta) = [\cos \theta, \cos 2\theta, \cdots, \cos m\theta],$$

and S is the region given by $\begin{vmatrix} 1 & \boldsymbol{I} \\ \boldsymbol{r} & \boldsymbol{X} \end{vmatrix} \geq 0$. This suggests, by analogy, that the joint distribution function is a polynomial in r_m of degree $2(\frac{1}{2}n - m - 1) + 3 = n - 2m + 1$ which vanishes only when $R_m = 0$. The equation satisfies these conditions, and in addition, it reduces to the known form when $m = 1$ and can be integrated to give this same form. Thus there is a strong suggestion that (13) gives an approximate distribution of $r_1, r_2, \cdots, r_m$, uncorrected for the mean.

An alternative form for the constant factor in (13) may be obtained if we note that

$$\frac{\Gamma(\frac{1}{2}n - m + 2)}{\Gamma(\frac{1}{2}n - m + \frac{3}{2})\pi^{\frac{1}{2}}} = \frac{1}{2^{n-2m+2}} \frac{\Gamma(n - 2m + 3)}{[\Gamma(\frac{1}{2}n - m + \frac{3}{2})]^2}.$$

d) Now r_p' and r_p'' can be written in the forms $(S_{p-1} + R_{p-1})/R_{p-2}$ and $(S_{p-1} - R_{p-1})/R_{p-2}$, where

$$S_{p-1} = (-1)^{p-1} \begin{vmatrix} r_1 & r_2 & r_3 & \cdots & 0 \\ 1 & r_1 & r_2 & \cdots & r_{p-1} \\ r_1 & 1 & r_1 & \cdots & r_{p-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{p-2} & r_{p-3} & r_{p-4} & \cdots & r_1 \end{vmatrix}.$$

Thus

$$R_p = R_{p-2}\left(\frac{S_{p-1} + R_{p-1}}{R_{p-2}} - r_p\right)\left(r_p - \frac{S_{p-1} - R_{p-1}}{R_{p-2}}\right)$$

$$= \frac{R_{p-1}^2}{R_{p-2}}\left[1 - \left(\frac{r_p R_{p-2} - S_{p-1}}{R_{p-1}}\right)^2\right]$$

$$Q_p = Q_{p-1}(1 - r_{1p+1.23\ldots}^2)$$

where $$r_{1,p+1.23\ldots} = T_{p-1}/R_{p-1},$$

and

$$T_{p-1} = \begin{vmatrix} r_p & r_1 & r_2 & \cdots & r_{p-1} \\ r_{p-1} & 1 & r_1 & \cdots & r_{p-2} \\ r_{p-2} & r_1 & 1 & \cdots & r_{p-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_1 & r_{p-2} & r_{p-3} & \cdots & 1 \end{vmatrix}.$$

Therefore, if we make a change of variable to $r_{1,p-1.23\cdots}$, $r_{1,p.23\cdots}$, $\cdots$ $r_{13.2}$, r_1, we find that the new variables which correspond exactly to partial correlation coefficients are, in fact, independently distributed as such, with 3 degrees of freedom more than in the case where the sets of variables are distinct observations.

While the above properties do not prove that the r_i or $\bar{r}_i$ may be tested using partial or multiple correlation coefficients, this conjecture has been verified elsewhere and it has been shown [8] that, with certain adjustments, a test can be derived which is applicable to fairly short series.

REFERENCES

[1] R. L. Anderson, "Distribution of the serial correlation coefficient," *Annals of Math. Stat.*, Vol. 13 (1942), pp. 1–13.

[2] T. Koopmans, "Serial correlation and quadratic forms in normal variables," *Annals of Math. Stat.*, Vol. 13 (1943), pp. 14–33.

[3] M. S. Bartlett, "On the theoretical specification of sampling properties of autocorrelated time series," *Roy. Stat. Soc. Suppl.*, Vol. 8 (1946), pp. 27–41.

[4] W. G. Madow, "Note on the distribution of the serial correlation coefficient," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 308–310.

[5] L. Garding, *Proceedings of Lund University Mathematical Seminars*, Vol. 5, pp. 185–202.

[6] W. J. Dixon, "Further contributions to the problem of serial correlation," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 119–144.

[7] M. H. Quenouille, "Some results in the testing of the serial correlation coefficient," *Biometrika*, Vol. 35 (1948), pp. 261–7.

[8] M. H. Quenouille, "Approximate tests of correlation in time series 1," *Roy. Stat. Soc. Suppl.*, Vol. 11 (1949).

ON THE ESTIMATION OF THE NUMBER OF CLASSES IN A POPULATION[1]

By Leo A. Goodman

Princeton University

1. Summary. This paper deals with the following problem: Suppose a population of known size N is subdivided into an unknown number of mutually exclusive classes. It is assumed that the class in which an element is contained may be determined, but that the classes are not ordered. Let us draw a random sample of n elements without replacement from the population. The problem is to estimate the total number K of classes which subdivide the population on the basis of the sample results and our knowledge of the population size.

There is exactly one real valued statistic S which is an unbiased estimate of K when the sample size n is not less than the maximum number q of elements contained in any class. The restriction placed upon q is unimportant for many practical problems where either there is a reasonably low bound for q or those classes containing more than n elements are known. An unbiased estimate does not exist when there is no such knowledge.

Since the unbiased estimate can be very unreasonable, modifications of S are considered. The statistic

$$T' = \begin{cases} S' = N - \dfrac{N(n-1)}{n(n-1)} x_2, & \text{if } S' \geq \sum_{i=1}^{n} x_i, \\ \sum_{i=1}^{n} x_i, & \text{if } S' < \sum_{i=1}^{n} x_i, \end{cases}$$

where x_i is the number of classes containing i elements in the sample, is the most suitable estimate, in comparison with three other statistics, for a hypothetical population.

The case where each element in the population has an equal and independent chance of coming into the sample is used as a model for some sampling procedures and also as an approximation to the case of random sampling.

2. Introduction. The problem discussed may be described in terms of colored balls in an urn. How should we estimate the number of colors present in the urn on the basis of both the sample which gives the number of, say, white balls, red balls, etc., and our knowledge of the total number of balls in the urn:

The following practical cases illustrate some of the ways in which this problem presents itself:

(1) A company has received a large number of requests for a free sample of its product. It is known that the same people often send more than one request.

---

[1] Prepared in connection with research sponsored by the Office of Naval Research.

From a sample of the requests we wish to estimate how many different people have sent requests.[2]

(2) The Social Security Board possesses a large collection of Social Security cards. It is known that some people obtain different cards when they change jobs. From a sample of the cards it is desired to estimate how many different people have Social Security cards.[3]

(3) A person who sells durable commodities anticipates opening a store which is to be located at a highway intersection. He would like to know how many different automobiles pass through the intersection in a given time period. The total number of automobiles may be easily observed but some probably pass through more than once. This type of inquiry is also useful to advertising agencies which must decide the most efficient location for billboards.

(4) The State Unemployment Compensation Board possesses a large list of the people receiving unemployment benefits. It is desired to estimate the total number of families benefiting from the insurance program on the basis of a random sample of the people named on the list.

(5) The number of words in a book may be easily estimated and a sample can be taken. The problem of estimating the number of different words in a book is another analogue of the general problem.[4]

3. Results and derivations. In order to show that an unbiased estimate of the number of classes in a population exists when the sample size n is not less than the maximum number q of elements contained in any class, we need prove the following two statements:

LEMMA 1. *Suppose we have K classes of N similar elements with n_1 elements in class 1, n_2 elements in class 2, $\cdots$, n_K elements in class K. The class of an element is readily identifiable when the element is examined. Let*

$$q = \max (n_s).$$

Suppose a random sample is drawn without replacement. If x_i is the number of classes containing i elements in the sample, and K_j is the number of classes containing j elements in the population, then

$$E(x_i) = \sum_{j=i}^{q} \Pr(i \mid j, N, n)K_j,$$

where $Pr(i \mid j, N, n)$ shall henceforth be an abbreviation of

$$\frac{C_i^j\, C_{n-i}^{N-j}}{C_n^N}.$$

[2] Submitted by Charles Callard to Question and Answers, *The American Statistician*, Vol. 3, No. 1, p. 23.

[3] Mentioned to the author by Dr. J. Stevens Stock of Opinion Research Corporation.

[4] Mentioned in letter to the author from Frederick Mosteller of Harvard University.

PROOF. Let y_s be the number of elements appearing in the sample from the s-th class. The statement is proved by considering $E(x_i) = \sum_{s=1}^{K} E(\delta_{iy_s})$, where

$$\delta_{iy_s} = \begin{cases} 1, & \text{if } y_s = i, \\ 0, & \text{if } y_s \neq i. \end{cases}$$

LEMMA 2. *Let*

$$a^{(t)} = \begin{cases} a(a-1)(a-2)\cdots(a-t+1), & \text{for } t > 0, \\ 1 & , \text{ for } t = 0. \end{cases}$$

If

$$a_i = 1 - (-1)^i \frac{[N-n+i-1]^{(i)}}{n^{(i)}},$$

then

$$\sum_{i=1}^{j} A_i \Pr(i \mid j, N, n) = 1.\text{[5]}$$

This result follows directly from the fact that

$$\sum_{i=0}^{j} (-1)^i C_i^j [N-n+i-1]^{(j-1)} = 0, \text{ for } j \geqq 1.$$

The following theorem may be proved directly by the preceding lemmas:

THEOREM 1. *Suppose a sample of n elements is drawn without replacement from a population of size N which is subdivided into K classes. Let*

$$A_i = 1 - (-1)^i \frac{[N-n+i-1]^{(i)}}{n^{(i)}}.$$

If there are x_i classes containing i elements in the sample, then

$$E\left(\sum_{i=1}^{n} A_i x_i\right) = K,$$

provided that n is not less than the maximum number q of elements contained in any class in the population.

THEOREM 2. *There is at most one real valued statistic which is an unbiased estimate of the number of classes in a population.*[6]

PROOF. Let us order the points of the sample space in the following manner: Letting x_i be the number of classes containing i elements in the sample, order the sample points by increasing values of x_n; for equal values of x_n, order the points by increasing values of x_{n-1}; for equal values of x_{n-1}, order the points

[5] The author is indebted to Professor Frederick F. Stephan of Princeton University for a statement leading to a simplification of the original result.

[6] This statement was mentioned to the author by M. P. Peisakoff of Princeton University.

by increasing values of x_{n-2} ; $\cdots$; for equal values of x_3, order the points by increasing x_2. Let

$$x_1 = n - \sum_{j=2}^{n} jx_j .$$

To prove the theorem, we must show that to each 0_i there corresponds a unique value $S(i)$, which must be the value of our estimate when 0_i is observed, in order that the statistic be unbiased. To each

$$0_i = [x_1(i), x_2(i), x_3(i), \cdots , x_n(i)],$$

let us associate the population

$$P_i = \left[N - \sum_{j=2}^{n} jx_j(i), x_2(i), x_3(i) \cdots , x_n(i)\right].$$

If P_1 is the underlying population, then 0_i for all $i > 1$ will occur with a probability of zero. Since there are N classes in P_1, the value of the statistic must be $S(1) = N$ whenever 0_1 is observed in order that the estimate be unbiased. The theorem may now be proved by induction.

Since all the P_i used in the proof of Theorem 2 satisfied the condition that the maximum number q of elements contained in any class be not more than the sample size n, the statistic S is the only real valued statistic which is an unbiased estimate when $q \leq n$.

When the restriction that $q \leq n$ is removed, it is useless to search for an unbiased estimate since we have

THEOREM 3. *There does not exist an unbiased estimate of the number of classes subdividing a population when it is not known whether the maximum number q of elements contained in any class is not more than the sample size n.*

By the preceding theorems it is clear that if an unbiased estimate exists it must equal S. However, S is generally not unbiased when $n < q$.

THEOREM 4. *Suppose the statistics S_1, S_2, $\cdots$, S_n are the solutions of the system of linear equations*

$$x_i = \sum_{j=1}^{n} \Pr(i \mid j, N, n)S_j , \text{ for } i = 1, 2, \cdots , n,$$

where x_i is the number of classes containing i elements in a sample of size n from a population of N elements. If K_j is the number of classes containing j elements in the population, then $E(S_j) = K_j$, for $j = 1, 2, \cdots , n$ when n is not less than the maximum number q of elements contained in any class.

PROOF. We observe that the statement is certainly true for $j = q + 1, q + 2, \cdots , n$, since

$$E(S_j) = K_j = 0, \quad \text{for} \quad j = q + 1, q + 2, \cdots , n.$$

The statement is also true for $j = q$, since

$$E(S_q) = E(x_q) \frac{N^{(q)}}{n^{(q)}} = K_q .$$

To prove that $E(S_j) = K_j$, for any $j < q$, we assume it to be true for all $i > j$, whereupon its truth for j follows.

By Theorem 2, and 3, it is clear that $\sum_{j=1}^{n} S_j = S$. Since

$$\sum_{j=1}^{q} jK_j = N,$$

it seems reasonable to ask whether the values of the estimates $S_1, S_2, \cdots, S_n$ are in agreement with the known value of the size of the population. The unbiased estimate of K can be shown to be internally consistent by

Theorem 5. *Suppose a sample of size n is drawn without replacement from a population of N elements which is divided into classes. If x_i is the number of classes containing i elements in the sample, and if the linear equations*

$$x_i = \sum_{j=1}^{n} \Pr(i \mid j, N, n) S_j,$$

are solved simultaneously for S_j, then

$$\sum_{j=1}^{n} jS_j = N.$$

The theorem follows readily from the fact that

$$\sum_{i=1}^{j} i \Pr(i \mid j, N, n) = n \frac{j}{N} \text{ and } \sum_{i=1}^{n} ix_i = n.$$

The variance of S may now be calculated by means of the formula

$$\sigma_S^2 = \sum_{i,j=1}^{n} A_i A_j u_{ij} = \sum_{i,j=1}^{n} A_i A_j \left\{ \sum_{s,t=1}^{q} m_{st}(i, j) K_s K_t \right.$$
$$\left. + \sum_{s=1}^{q} [m_s(i, j) - m_{ss}(i, j)] K_s \right\},$$

where u_{ij} is the covariance between x_i and x_j, $m_{st}(i, j)$ is the covariance between δ_{iy_r} and δ_{jy_h} when $r \neq h$, $n_r = s$ and $n_h = t$, and $m_s(i, j)$ is the covariance between δ_{iy_r} and δ_{jy_r} when $n_r = s$.

Since the statistic S can be very unreasonable, we consider other possible estimates of K. The statistic

$$S' = N - \frac{N^{(2)}}{n^{(2)}} x_2$$

may be shown to be a modification of S which replaces the number x_i of classes containing $i > 2$ elements in the sample by an additional ix_i classes, each containing only one element. Since the values of K_i for $i > 2$ are relatively small in the practical problems of Section 2, S' might be used as an estimate.

Another statistic which may be used to estimate K is

$$S'' = \frac{N}{n} \sum_{i=1}^{n} x_i.$$

This statistic may be shown to overestimate K whenever $q \neq 1$. The estimate

$$S'' = \sum_{i=1}^{n} x_i$$

underestimates K when $n \leq N - m$ where m is the least number of elements contained in any class.

4. Binomial sampling. Let us suppose that each element from a population of N elements has an equal and independent chance $p = 1/r$ of entering the sample s. In this case, the size of the sample obtained is a random variable η which is binomially distributed with mean Np. If a large random sample of n elements is drawn without replacement from a large population of size N, then the results when interpreted in terms of binomial samples where $p = 1/r = n/N$ are a good approximation to the results obtained by the usual model. Binomial sampling may be considered a model of the case where one attempts to obtain the sampling ratio $p = 1/r$ by drawing simultaneously an uncounted sample of elements which is estimated as being of the appropriate size.

In the case of binomial sampling, the statistic

$$B = \sum_{i=1}^{N} B_i x_i, \quad \text{where} \quad B_i = 1 - (1 - r)^i$$

may be shown to be an unbiased estimate of the number of classes in a population from which binomial samples are drawn.

Let us now consider the statistic which corresponds to S' for the case of binomial sampling; i.e.,

$$B' = N - r^2 x_2 .$$

It may be shown that

$$E(B') = K_1 + K_2 + \sum_{j=3}^{q} [j - C_2^j (1 - p)^{j-2}] K_j .$$

Hence, the statistic B' will underestimate K whenever

$$p < 1 - \left(\frac{2}{j}\right)^{1/j-2}, \quad \text{for} \quad j = 3, 4, \cdots, q.$$

Since

$$1 - \left(\frac{2}{j}\right)^{1/j-2}$$

is a decreasing function of j for $j > 2$, when $p > \frac{1}{3}$, B' overestimates, and when

$$p < 1 - \left(\frac{2}{p}\right)^{1/q-2},$$

B' underestimates the value of K. When p is such that

$$1 - \left(\frac{2}{q}\right)^{1/q-2} \leq p \leq \tfrac{1}{3},$$

the expected value of B' is brought closer to K by underweighting some K_j and overweighting others.

5. A hypothetical population.[7] Suppose we draw a random sample of 1000 elements without replacement from a population of 10,000 elements where

$$K_1 = 9225, \qquad K_2 = 336, \qquad K_3 = 33, \qquad K_4 = 1.$$

Hence, $K = 9595$. By means of Table 1, let us now compare on the basis of binomial sampling the estimates which have been presented in the preceding sections. Since N and n are large, these results are a good approximation to the case of random sampling without replacement.

TABLE 1

Estimate	*Expected value*	*Bias*	$\sqrt{\textit{Mean Square Error}}$
S	9595	0	347
S'	9570	−25	207
S''	9959	364	490
S'''	996	−8599	8600

It is clear that the best estimates of the number of classes in this particular population are S or S', since S has the least bias, $E(S) - K$, and S' has the least mean square error, $E(S' - K)^2$. One might argue that both S and S' are the statistics which are capable of giving nonsensical estimates. However, we may decide to modify S or S' in order to always get reasonable estimates by using the statistics

$$T = \begin{cases} S, & \text{if } N \geq S \geq \sum_{i=1}^{n} x_i, \\ N, & \text{if } S > N, \\ \sum_{i=1}^{n} x_i, & \text{if } S < \sum_{i=1}^{n} x_i \end{cases}$$

$$T' = \begin{cases} S', & \text{if } S' \geq \sum_{i=1}^{n} x_i, \\ \sum_{i=1}^{n} x_i, & \text{if } S' < \sum_{i=1}^{n} x_i. \end{cases}$$

[7] Other examples have been investigated by Frederick Mosteller in Questions and Answers, *The American Statistician*, Vol. 3, No. 3, p. 12.

Although these modified statistics T and T' are not unbiased, they have the desirable property that

$$MSE(T) \leq MSE(S), \quad \text{and} \quad MSE(T') \leq MSE(S').$$

Since this hypothetical population is a plausible one for the practical problems of Section 2, the modified statistics T or T' seem, therefore, to be "best" for estimating the number of classes for these problems, where the "best" statistic is defined as the one which never gives unreasonable estimates and has the least mean square error.

The author wishes to express his appreciation to Professor John W. Tukey whose suggestions were very helpful.

CONCERNING COMPOUND RANDOMIZATION IN THE BINARY SYSTEM

By John E. Walsh

The Rand Corporation

1. Summary. Let us consider a set of approximately random binary digits obtained by some experimental process. This paper outlines a method of compounding the digits of this set to obtain a smaller set of binary digits which is much more nearly random. The method presented has the property that the number of digits in the compounded set is a reasonably large fraction (say of the magnitude $\frac{1}{3}$ or $\frac{1}{4}$) of the original number of digits.

If a set of very nearly random decimal digits is required, this can be obtained by first finding a set of very nearly random binary digits and then converting these digits to decimal digits.

The concept of "maximum bias" is introduced to measure the degree of randomness of a set of digits. A small maximum bias shows that the set is very nearly random.

The question of when a table of approximately random digits can be considered suitable for use as a random digit table is investigated. It is found that a table will be satisfactory for the usual types of situations to which a random digit table is applied if the reciprocal of the number of digits in the table is noticeably greater than the maximum bias of the table.

2. Introduction and discussion. With the development of the theory of games and the more widespread use of experimental methods for determining approximate distributions for statistics whose probability laws are difficult to obtain analytically, a demand for large sets of random digits has arisen. The problem of obtaining a set of digits which can be considered sufficiently random for the situations to which it would be applied, however, is not an easy one. One approach to this problem consists in obtaining a set of digits by some procedure and then applying tests to this set of digits to determine whether it can be considered satisfactory. Although appropriate choice of the tests may result in acceptance of sets of digits which are suitable for certain special types of situations, this approach is of a negative character and does not prove that a given set of digits is sufficiently random; it merely indicates that this may be the case. What is needed is a constructive approach to the problem, i.e., a method of constructing a set of random digits which can be proved sufficiently random for most applications if certain intuitively acceptable conditions are satisfied. A step in this direction has already been taken by H. Burke Horton in [1] and by H. Burke Horton and R. Tynes Smith III in [2]. This paper presents what is hoped will be another step in this direction.

In this paper, considerations will be limited to the case of binary digits. The reasons for this are twofold:

(a). The method used for compounding the digits yields a sharp upper bound for the maximum bias of the compounded set (i.e., a bound that the maximum bias could actually attain) only for the case of binary digits.

(b). Many of the experimental procedures for obtaining approximately random digits consist in first producing binary digits and then converting to another number base. Thus binary digits are produced directly. Hence, to use the results of this paper, the only modification required in these procedures would be to compound the binary digits before they are converted.

Now let us consider some definitions: A set of random variables each of which can assume only the values 0 and 1 will be referred to as a set of binary digits. For convenience, each of the random variables making up a set of binary digits will be called a binary digit; this is not to be confused with the value obtained for the random variable. The absolute value of the deviation from $\frac{1}{2}$ of the conditional probability that a specified binary digit has the value 0 (or 1) is called the *bias* of that digit for the given conditions on the remaining digits of the set. The maximum bias of a binary digit is defined to be the maximum of the biases of that digit with respect to all possible conditions on the remaining digits of the set. The *maximum bias of the set* is the greatest of the maximum biases of the digits of the set. A set of binary digits is said to be *random* if its maximum bias is zero.

The method used to prove that a set of compounded digits has a sufficiently small maximum bias is somewhat similar to the situation encountered in mathematics where one begins with certain axioms and then draws conclusions. If the axioms are correct, the conclusions are necessarily valid. The first step in the compounding procedure consists in obtaining a set of binary digits by some experimental process (perhaps from a random digit machine which is based on some physical principle). The experimental process is so chosen that there is no doubt that the set of binary digits produced satisfies the two conditions:

(i). The maximum bias of the set is less than or equal to some specified value $\alpha(<\frac{1}{2})$.

(ii). The digits of the set can be arranged in a specified array which has the property that the rows of the array are statistically independent.

On the basis of these two assumptions (which play the same role as the axioms mentioned above), it can be proved that the maximum bias of the resulting compounded set of binary digits never exceeds a specified value which depends on α. Moreover, the upper bound for the maximum bias of the constructed set of binary digits can be made extremely small even for large values of α.

If the experimental process is suitably chosen, conditions (i) and (ii) can be satisfied beyond any doubt. For example, let us consider 1000 people located in different parts of the world and not in contact with each other. Let each person flip an ordinary coin high in the air so that it will land on a flat hard surface, record the result (say 0 for a tail and 1 for a head), and then repeat this procedure until 5000 binary digits are obtained. If α is set equal to 3/10, condition (i) is

obviously satisfied for the resulting set of 5,000,000 binary digits. Condition (ii) evidently holds if the array is taken to consist of 1000 rows where each row contains 5000 binary digits obtained from one person.

The ideal choice for α would be the actual maximum bias of the set of binary digits obtained from the experimental process. Then the compounding procedure for obtaining a set of digits with a specified upper bound for the maximum bias would be simplified; also the number of digits in the compounded set would be a larger fraction of the original number of digits. Invariably, however, the properties of the experimental process are not known with sufficient accuracy for obtaining anything but a safe upper bound on the maximum bias of the set of digits produced. This situation is analogous to that of estimating the length of a stick which a very rough measurement has shown to be about 10″ long. Although one might be very hesitant to believe that the length of the stick lies between 9.9″ and 10.1″, the contention that the length lies between 5″ and 15″ can be accepted with virtual certainty and any logical conclusions based on this contention can also be accepted with virtual certainty.

Given the number of binary digits in a set and the maximum bias of the set, is it possible to determine whether the set is suitable for use as a set of random binary digits? An important consideration in answering this question is the use that is to be made of the set of digits. This must always be taken into account before the suitability of the set can be decided. For example, if no more than 1/1000 of the digits of the set are to be used for any particular situation, the set might be satisfactory for the types of cases to which it would be applied; on the other hand, the set might not be suitable for cases of these types if all the digits of the set are used for each situation. This example calls attention to an important point, namely that the suitability of a set of binary digits depends on the number of digits in the set. Let a set have a fixed non-zero maximum bias ρ. If the set contains a sufficiently large number N of digits, relations and expressions involving the digits of the set can be found whose probabilities, moments, etc., can differ greatly from the values which would be obtained if the relations were based on the same number of truly random binary digits. As a specific example consider the relation

All the digits of the set have the value zero.

If the reciprocal of the number of digits in the set is of the same order of magnitude or smaller than the maximum bias of the set, the ratio of the probability of this expression to its hypothetical value can differ noticeably from unity. Thus, at least in certain special cases, a necessary condition for the suitability of a set of binary digits is that $1/N >> \rho$. This condition, however, is also sufficient for most situations to which a set of random digits would be applied. The approximate sufficiency of the condition is a direct consequence of the fact that any set of N binary digits can be considered as a sample value from an N-dimensional population consisting of 2^N discrete points. The $1/N >> \rho$ restriction implies that the probability concentrated at each of the 2^N points is

very nearly equal to the hypothetical value of $(\frac{1}{2})^N$ for all possible conditions on the remaining digits of the set.

The $1/N >> \rho$ condition is very satisfactory from the viewpoint of probabilities. The probability of any relation based on a subset of the digits of the set (possibly conditioned on other digits from the table) can be interpreted as the sum of the probabilities of those points included in a certain region (defined by the relation) of the N-dimensional probability space of the set of digits. By expanding $(\frac{1}{2} \pm \rho)^N$ it can be shown that the ratio of the probability of any relation based on one or more digits from the set to the corresponding value for a truly random set of digits will be very nearly equal to unity if $1/N >> \rho$.

It is evident that the higher order moments of an expression based on one or more digits of the set can differ noticeably from its hypothetical value even if $1/N >> \rho$; any deviation from the ideal situation, no matter how small, can become important for high order moments. For the first few moments, however, deviations from the hypothetical values are not appreciable since these moments are based on the probabilities at the 2^N points in the N-dimensional probability space and these probabilities are very nearly equal to the hypothetical value of $(\frac{1}{2})^N$ in all cases.

The above discussion shows that the values of N and ρ are sufficient to determine whether a set of binary digits is suitable for use as random binary digits for a wide variety of situations. Analogous considerations apply for digits to any number base.

A magnitude definition of the relation $1/N >> \rho$ is difficult to specify. If ρ is the upper bound for the maximum bias of a set of digits obtained by the compounding procedure outlined in this paper, however, it seems that a reasonable condition would be that $1/N \geq 50\rho$. This condition implies that the probability of any relation based on digits of the set can not differ from its hypothetical value by more than approximately 4%. In most practical applications the value obtained for ρ would be noticeably greater than the true value of the maximum bias of the compounded set.

Since the maximum number of digits which can be taken from a table is the total number of digits in the table, the above considerations suggest that a random digit table should be constructed so that the reciprocal of the number of digits in the table is noticeably greater than the maximum bias of the table. Any table having this property would be satisfactory for most situations to which it would be applied.

Now let us consider two different compounding methods which produce sets of binary digits with the same upper bound for the maximum bias. If the computational difficulties of applying the two methods are of comparable magnitudes, it seems reasonable to prefer the method which yields the larger set of digits. For example, if the number of digits in the set obtained by the first method is only 1/8 of the original number of digits while the number in the set obtained by the second method is 1/3 of the original number, the second method would seem preferable even if it required as much as 100% more computation.

The compounding method presented in this paper has the property that the number of digits in the compounded set can be held to a reasonably large fraction of the original number of digits at the same time that the upper bound for the maximum bias is made extremely small. The method presented by Horton in [1] does not have this property. For example, let $\alpha = 1/10$. Applying Horton's method, when the compounded set consists of 1/8 of the original number of digits the upper bound for the maximum bias is 12.8×10^{-7}. The example presented in section 3, however, shows that a compounded set whose number of digits equals 1/3 of the original number and which has an upper limit of 11.7×10^{-7} for the maximum bias can be obtained using the method presented in the next section.

Although the compounding method outlined in section 3 is presented as a series of steps, the value of a digit of the compounded set can be written as a linear function (mod 2) of digits of the original set. This was not done in what follows because of the complicated nature of the general form of such expressions. In any particular case, however, these expressions can be written without much trouble and the compounded digits computed from the original digits in a single step.

3. Outline of compounding method and statement of theorems. This section contains a description of the compounding method mentioned in the preceding two sections as well as statements of the basic theorems concerning this compounding method. Proofs of the results stated in this section are given in section 4.

Let us consider the array of mn binary digits

$$\begin{matrix} x_{11}, & x_{12}, & \cdots, & x_{1n} \\ x_{21}, & x_{22}, & \cdots, & x_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ x_{m1}, & x_{m2}, & \cdots, & x_{mn} \end{matrix} \tag{1}$$

which satisfies conditions (i) and (ii); i.e., the maximum bias of the set (1) is less than or equal to α while a digit x_{uv} is independent of a digit x_{rs} if $r \neq u$ (if $r = u$, however, x_{uv} is not necessarily independent of x_{rs}).

Let a new set of $(m - 1)n$ binary digits

$$y_{ij}, \qquad (i = 1, \cdots, m-1; j = 1, \cdots, n) \tag{2}$$

be formed as follows:

$$y_{ij} = x_{mj} + x_{ij} \pmod 2,$$
$$(i = 1, \cdots, m-1; j = 1, \cdots, n).$$

Then the biases of the y_{ij} have the properties

THEOREM 1. *Let U be a specified set of $t - 1$ of $y_{1j}, \cdots, y_{(i-1)j}, y_{(i+1)j}, \cdots, y_{(m-1)j}$, $(1 \leq t \leq m - 1)$, while V is a specified set of zero or more of the y_{pq}'s*

with $q \neq j$. Also let θ consist of the set of integers such that $p \in \theta$ if $y_{pj} \in U$. Then, if γ_u = maximum bias for the set $x_{u1}, \cdots, x_{un}$, $(u = 1, \cdots, n)$,

$$|Pr(y_{ij} = 0 \mid U, V) - \tfrac{1}{2}| \leq \gamma_i \left[1 - \prod_{k \in \theta} (\tfrac{1}{2} - \gamma_k)/(\tfrac{1}{2} + \gamma_k)\right] \Big/ \left[1 + \prod_{k \in \theta} (\tfrac{1}{2} - \gamma_k)/(\tfrac{1}{2} + \gamma_k)\right]$$

for all possible selections of U, V and of the values for the digits of these sets.

COROLLARY 1. *If exactly $t - 1$ of $y_{1j}, \cdots, y_{(i-1)j}, y_{(i+1)j}, \cdots, y_{(m-1)j}$ have known values, the maximum bias of the binary digit y_{ij} is less than or equal to*

$$\alpha[1 - (\tfrac{1}{2} - \alpha)^t/(\tfrac{1}{2} + \alpha)^t]/[1 + (\tfrac{1}{2} - \alpha)^t/(\tfrac{1}{2} + \alpha)^t].$$

COROLLARY 2. *The maximum bias of the set (2) is less than or equal to*

$$\alpha[1 - (\tfrac{1}{2} - \alpha)^{m-1}/(\tfrac{1}{2} + \alpha)^{m-1}]/[1 + (\tfrac{1}{2} - \alpha)^{m-1}/(\tfrac{1}{2} + \alpha)^{m-1}].$$

The basic operation in the method of compounding binary digits is outlined in the procedure given for obtaining the y_{ij} from the x_{uv}. Let $m = (1 + t_1) \cdots (1 + t_K)$. Then a set of $t_1 \cdots t_K n$ binary digits can be obtained from the original set of mn digits x_{uv} by continually applying this basic procedure. The first step consists in dividing the rows of (1) into $(1 + t_2) \cdots (1 + t_K)$ sets each consisting of $(1 + t_1)$ rows in some specified fashion. Each of these sets is an array of $(1 + t_1) \times n$ binary digits for which the rows are independent. Apply the method used to obtain the y_{ij} from the x_{uv} to each $(1 + t_1) \times n$ array separately. Then each array yields a set of $t_1 n$ binary digits and there are $(1 + t_2) \cdots (1 + t_K)$ such sets. In each set arrange the $t_1 n$ digits into a single row in some specified manner. This furnishes a new array of $[(1 + t_2) \cdots (1 + t_K)] \times [t_1 n]$ binary digits for which the rows are independent. Repeat this procedure with respect to t_2 thus obtaining a new array of $[(1 + t_3) \cdots (1 + t_K)] \times [t_1 t_2 n]$ binary digits for which the rows are independent; etc., until a $(1 + t_K) \times (t_1 \cdots t_{K-1} n)$ binary digit array for which the rows are independent is obtained. Then form a set of binary digits Y_{gh}, $(g = 1, \cdots, t_K; h = 1, \cdots, t_1 \cdots t_{K-1} n)$, from this array in exactly the same manner that the y_{ij} were obtained from the x_{uv}. Then the biases of the Y_{gh} have the properties

THEOREM 2. *Let $\beta_0, \beta_1, \cdots, \beta_K$ be defined by $\beta_0 = \alpha$ and*

$$\beta_w = \beta_{w-1}[1 - (\tfrac{1}{2} - \beta_{w-1})^{t_w}/(\tfrac{1}{2} + \beta_{w-1})^{t_w}]/[1 + (\tfrac{1}{2} - \beta_{w-1})^{t_w}/(\tfrac{1}{2} + \beta_{w-1})^{t_w}], \quad (w = 1, \cdots, K).$$

Then, if exactly $t - 1$ of $Y_{1h}, \cdots, Y_{(g-1)h}, Y_{(g+1)h}, \cdots, Y_{t_K h}$ have known values, $(1 \leq t \leq t_K)$, the maximum bias of the digit Y_{gh} is less than or equal to

$$\beta_{K-1}[1 - (\tfrac{1}{2} - \beta_{K-1})^t/(\tfrac{1}{2} + \beta_{K-1})^t]/[1 + (\tfrac{1}{2} - \beta_{K-1})^t/(\tfrac{1}{2} + \beta_{K-1})^t].$$

In particular, the maximum bias of the entire set of Y_{gh} is less than or equal to β_K. Also

$$\beta_{K-1}[1 - (\tfrac{1}{2} - \beta_{K-1})^t/(\tfrac{1}{2} + \beta_{K-1})^t]/[1 + (\tfrac{1}{2} - \beta_{K-1})^t/(\tfrac{1}{2} + \beta_{K-1})^t] \leq 2^{2^{K}-1} \cdot t \cdot t_{K-1}^{2} \cdot t_{K-2}^{4} \cdots t_2^{2^{K-2}} \cdot t_1^{2^{K-1}} \cdot \alpha^{2^K}. \tag{3}$$

The inequality (3) is frequently useful from a computational viewpoint. Although the right hand side of (3) is usually noticeably greater than the left hand side, in many cases this rough upper bound is itself small enough to show that the upper bound for the maximum bias is of the desired order of magnitude.

If the set of compounded digits is to be used for a random binary digit table, Theorem 2 shows that advantage can be taken of the position of the digits in the table. Let $M = t_1 \cdots t_{K-1}n$ and enter the values of the Y_{gh}, $(g = 1, \cdots, t_K;$ $h = 1, \cdots, M)$, into the table in the order

$$Y_{11}, Y_{12}, \cdots, Y_{1M}, Y_{21}, \cdots, Y_{2M}, Y_{31}, \cdots, Y_{t_K 1}, \cdots, Y_{t_K M}.$$

Then, if a set of digits is taken from this table in consecutive order (Y_{11} follows $Y_{t_K M}$), the upper bound for the maximum bias of this set is dependent on the number L of digits in the set. From Theorem 2, the maximum bias of a set of L digits taken in consecutive order from a table formed in this manner is less than or equal to

$$\beta_{K-1}[1 - (\tfrac{1}{2} - \beta_{K-1})^t/(\tfrac{1}{2} + \beta_{K-1})^t]/[1 + (\tfrac{1}{2} - \beta_{K-1})^t/(\tfrac{1}{2} + \beta_{K-1})^t]$$

for values of L such that $(t - 1)M < L \leq tM$, where $1 \leq t \leq t_K$. Thus, if a small set of digits is taken from this table in consecutive order, the upper bound for the maximum bias of this set will usually be noticeably smaller than the upper bound for the maximum bias of the table. Since many uses of a random digit table require only a small fraction of the total number of entries in this table, this property would seem to be desirable. It should be emphasized, however, that the maximum bias of a set taken from this table is always less than or equal to β_K irrespective of the positions that the digits of the sets occupy in the table. Thus nothing is lost by constructing the table in this manner but something can be gained for small sets if the digits are taken from the table in consecutive order.

Now let us consider situations in which it is required that the number of digits in the compounded set is at least a specified fraction, say $1/C$, of the original number mn of binary digits. This requires that K and $t_1, \cdots, t_K$ be chosen so that

$$t_1 \cdots t_K/(1 + t_1) \cdots (1 + t_K) \geq 1/C.$$

Also, for given values of K and C, it seems preferable to choose $t_1, \cdots, t_K$ so that the value of β_K is at least approximately minimized. Examination of the results of Theorem 2 indicates that a reasonable method of determining the values of $t_1, \cdots, t_K$ with this in mind consists in first choosing t_1 as small as possible, then (given the value of t_1 equal to its minimum value) choosing t_2 as small as possible, etc. This method is also recommended by the fact that the resulting values of $t_1, \cdots, t_K$ are readily determined. The explicit procedure for finding $t_1, \cdots, t_K$ is given by

THEOREM 3. *Let the values of the integer K and the constant C (> 1) be given and consider the integers $t_1, \cdots, t_K$ subject to the condition*

$$t_1 \cdots t_K/(1 + t_1) \cdots (1 + t_K) \geq 1/C.$$

The minimum value of t_1 is the smallest integer satisfying

$$t_1 > 1/(C - 1).$$

In general, $2 \leq w \leq K - 1$, having already determined $t_1, \cdots, t_{w-1}$ as their minimum values, the value of t_w is the smallest integer satisfying

$$t_w > 1/[Ct_1 \cdots t_{w-1}/(1 + t_1) \cdots (1 + t_{w-1}) - 1].$$

Finally, given $t_1, \cdots, t_{K-1}$ as their minimum values, the minimum value of t_K is the smallest integer satisfying

$$t_K \geq 1/[Ct_1 \cdots t_{K-1}/(1 + t_1) \cdots (1 + t_{K-1}) - 1].$$

Now consider the general situation encountered in the application of the compounding process outlined above. Here the values of α, C are given and it is required to choose K and $t_1, \cdots, t_K$ so that the upper bound for the maximum bias of the compounded set of $t_1 \cdots t_K n$ binary digits Y_{gh} is less than or equal to a specified value b. The following procedure furnishes a method of solving this problem:

Let $K = 1$, obtain t_1 according to Theorem 3, and then compute β_1. If $\beta_1 \leq b$, a solution has been obtained. If $\beta_1 > b$, let $K = 2$ and repeat the procedure to obtain β_2. If $\beta_2 \leq b$, the values of t_1, t_2 and $K = 2$ are a solution. If $\beta_2 > b$, repeat the procedure for $K = 3$; etc. In practical situations, the value of K is usually bounded (e.g., by independence properties of the original set of digits). If β_K is still greater than b for the maximum permissible value of K, no solution is obtained. This means that either b must be increased or $1/C$ decreased or both if a solution is to be found. In many cases, a large amount of computation can be avoided by using the inequality (3). For marginal situations, however, a solution may be missed by using (3) instead of computing β_K.

Example of method. The following table represents an example of application of the above method:

$\alpha = 1/10$	$1/C = 1/3$	$b = 2 \times 10^{-6}$
$K = 1,\ t_1 = 1$		$\beta_1 = 2 \times 10^{-2}$
$K = 2,\ t_1 = 1,\ t_2 = 2$		$\beta_2 \leq 1.6 \times 10^{-3}$
$K = 3,\ t_1 = 1, t_2 = 3,\ t_3 = 9$		$\beta_3 \leq 1.04 \times 10^{-4}$
$K = 4,\ t_1 = 1,\ t_2 = 3,\ t_3 = 10.\ t_4 = 44$		$\beta_4 \leq 1.17 \times 10^{-6}$.

Thus $K = 4, t_1 = 1, t_2 = 3, t_3 = 10, t_4 = 44$ is a solution.

4. Derivations. The purpose of this section is to furnish proofs of the results stated in the preceding sections.

4.1 *Proof of Theorem* 1. Let us consider the conditional probability that an arbitrary but fixed y_{ij} has a specified value when the values of a fixed subset of zero or more of the remaining y's are known. For convenience, assume that y_{11} is the binary digit considered and that the values of $y_{21}, y_{31}, \cdots, y_{t1}$ (where t is a fixed integer such that $1 \leq t \leq m - 1$) and a set S are given while the

values of the remaining y's are unknown. Here S represents an arbitrary but fixed set of zero or more of the y_{ij}'s for which $j \geq 2$ while $t = 1$ has the interpretation that none of the y_{i1}, $(i \geq 2)$, are given. Let

$$Pr(x_{m1} = 0 \mid S) = \tfrac{1}{2} + \alpha_{t+1} \quad \text{and} \quad Pr(x_{k1} = b_k \mid S) = \tfrac{1}{2} + \alpha_k, \qquad (k = 1, \cdots, t).$$

Then, using the independence conditions satisfied by the x's,

$$\begin{aligned} Pr(y_{11} &= b_1 \mid y_{21} = b_2, \cdots, y_{t1} = b_t; S) \\ &= \left[\prod_{k=1}^{t+1} (\tfrac{1}{2} + \alpha_k) + \prod_{k=1}^{t+1} (\tfrac{1}{2} - \alpha_k)\right] \Big/ \left[\prod_{k=2}^{t+1} (\tfrac{1}{2} + \alpha_k) + \prod_{k=2}^{t+1} (\tfrac{1}{2} - \alpha_k)\right] \\ &= \tfrac{1}{2} + \alpha_1 \left[\prod_{k=2}^{t+1} (\tfrac{1}{2} + \alpha_k) - \prod_{k=2}^{t+1} (\tfrac{1}{2} - \alpha_k)\right] \Big/ \left[\prod_{k=2}^{t+1} (\tfrac{1}{2} + \alpha_k) + \prod_{k=2}^{t+1} (\tfrac{1}{2} - \alpha_k)\right] \\ &= \tfrac{1}{2} + \alpha_1 \delta. \end{aligned}$$

Now $|\delta| = (1 - P)/(1 + P)$ if $0 \leq P \leq 1$ and equals $(P - 1)/(1 + P)$ if $P > 1$, where $P = \prod_{k=2}^{t+1} (\frac{1}{2} - \alpha_k)/(\frac{1}{2} + \alpha_k)$. Let γ_u be the maximum bias for the set of binary digits $x_{u1}, \cdots, x_{un}$, $(u = 1, \cdots, m)$. Then it is easily seen that

$$\max_P |\delta| \leq \left[1 - \prod_{k=2}^{t+1} (\tfrac{1}{2} - \gamma_k)/(\tfrac{1}{2} + \gamma_k)\right] \Big/ \left[1 + \prod_{k=2}^{t+1} (\tfrac{1}{2} - \gamma_k)/(\tfrac{1}{2} + \gamma_k)\right].$$

Thus

$$\begin{aligned} | Pr(y_{11} &= b_1 \mid y_{21} = b_2, \cdots, y_{t1} = b_t; S) - \tfrac{1}{2} | \\ &\leq \gamma_1 \left[1 - \prod_{k=2}^{t+1} (\tfrac{1}{2} - \gamma_k)/(\tfrac{1}{2} + \gamma_k)\right] \Big/ \left[1 + \prod_{k=2}^{t+1} (\tfrac{1}{2} - \gamma_k)/(\tfrac{1}{2} + \gamma_k)\right] \end{aligned}$$

for all possible selections of $b_1, \cdots, b_t$ and all possible selections of S and the values for the digits of S. It is to be observed that this inequality is valid for $t = 1$.

Evidently this result can be modified to apply to an arbitrary y_{ij} for which $t - 1$ of $y_{1j}, \cdots, y_{(i-1)j}, y_{(i+1)j}, \cdots, y_{(m-1)j}$ have given values. This obvious modification results in Theorem 1.

4.2 *Proof of Theorem* 2. By Corollary 2, the maximum bias of the $[(1 + t_2) \cdots (1 + t_K)] \times [t_1 n]$ array is less than or equal to β_1. In general, $2 \leq w \leq K$, by Corollary 2 the maximum bias of the $[(1 + t_{w+1}) \cdots (1 + t_K)] \times [t_1 \cdots t_{wn}]$ array is less than or equal to β_w. Finally, by Corollary 1, if exactly $t - 1$ of $Y_{1h}, \cdots, Y_{(g-1)h}, Y_{(g+1)h}, \cdots, Y_{t_K h}$ have known values, $(1 \leq t \leq t_K)$, the maximum bias for the binary digit Y_{gh} is less than or equal to

$$\beta_{K-1}[1 - (\tfrac{1}{2} - \beta_{K-1})^t/(\tfrac{1}{2} + \beta_{K-1})^t]/[1 + (\tfrac{1}{2} - \beta_{K-1})^t/(\tfrac{1}{2} + \beta_{K-1})^t].$$

The inequality (3) is an immediate consequence of the relation

$$\alpha[1 - (\tfrac{1}{2} - \alpha)^s/(\tfrac{1}{2} + \alpha)^s]/[1 + (\tfrac{1}{2} - \alpha)^s/(\tfrac{1}{2} + \alpha)^s] \leq 2s\alpha^2.$$

4.3 *Proof of Theorem* 3. From the given condition

$$t_K \geq 1/[Ct_1 \cdots t_{K-1}/(1 + t_1) \cdots (1 + t_{K-1}) - 1].$$

From this inequality for t_K it follows that

$$Ct_1 \cdots t_{K-1}/(1 + t_1) \cdots (1 + t_{K-1}) - 1 > 0.$$

Thus

$$t_{K-1} > 1/[Ct_1 \cdots t_{K-2}/(1 + t_1) \cdots (1 - t_{K-2}) - 1].$$

In general, $3 \leq w \leq K - 1$, given

$$t_w > 1/[Ct_1 \cdots t_{w-1}/(1 + t_1) \cdots (1 + t_{w-1}) - 1]$$

it follows that

$$Ct_1 \cdots t_{w-1}/(1 + t_1) \cdots (1 + t_{w-1}) - 1 > 0$$

whence

$$t_{w-1} > 1/[Ct_1 \cdots t_{w-2}/(1 + t_1) \cdots (1 + t_{w-2}) - 1].$$

Finally

$$t_1 > 1/(C - 1).$$

REFERENCES

[1] H. Burke Horton, "A method for obtaining random numbers," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 81–85.

[2] H. Burke Horton and R. Tynes Smith III, "A direct method for producing random digits in any number system," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 82–90.

THE DISTRIBUTION OF EXTREME VALUES IN SAMPLES WHOSE MEMBERS ARE SUBJECT TO A MARKOFF CHAIN CONDITION

By Benjamin Epstein

Department of Mathematics, Wayne University

1. Introduction. The extreme value problem as treated in the literature concerns itself with the following question: To find the distribution of the smallest, largest, or more generally the νth largest, or νth smallest values in random samples of size n, drawn from a distribution whose probability law is given by the d.f. $F(x)$. In this formulation the observed sample values $x_1, \cdots, x_n$ are assumed to be statistically independent. While the assumption of independence may be a good approximation to the true state of affairs in some cases, there are situations where this assumption is not justified.

Suppose, for instance, that the observations in the sample are ordered in time. Then it may happen that successive observations are stochastically dependent, the extent of this dependence being a function of the time interval separating these observations.[1] In such cases the present distribution theory for extreme values in samples of size n is inadequate and must be replaced by more general results.

It is clear that a clean-cut analytic solution to the problem of the distribution of extreme values in samples whose members may be stochastically dependent can be expected only for certain special kinds of dependence among successive observations. We are able, in this paper, to obtain the distribution of smallest, largest, second smallest, and second largest values in samples of size n drawn at equally spaced time intervals from a stationary Markoff process.

2. The distribution of smallest and largest values in samples of size n drawn at equally spaced time intervals from a stationary Markoff process. In this section the following assumption is made:

(A) observations $x_1, x_2, \cdots, x_n, \cdots$ are taken in order at times $t = 1$, $t = 2, \cdots, t = n, \cdots$ from a stationary Markoff random process.

The only information needed in the investigation of a stationary Markoff process at integral values of time is the function

$$F_2(x, y) = \text{Prob } (x_i \leq x, x_{i+1} \leq y), \tag{1}$$

independently of i, where $F_2(x, y)$ must be such that the marginal distribution obtained by integrating over x or y (if x_i or x_{i+1} take on a continuous range of

[1] If the observations $x_1, x_2, \cdots, x_n, \cdots$ are taken at discrete times $t_1, t_2, \cdots, t_n, \cdots$ a measure of stochastic dependence between x_i and x_j is the ordinary coefficient of correlation r_{ij}. If the observations are taken from a continuous stochastic process a natural measure of stochastic dependence between observations made at two different times is the covariance function of the process. In this paper we shall limit ourselves to processes which are discrete in time.

values) or summing over the possible values of x_i or x_{i+1} (if x_i and x_{i+1} can take on only discrete values) is of the form

$$F_1(x) = \text{Prob}\,(x_i \leq x), \tag{2}$$

independently of i.

An example of a random process meeting condition A is furnished by the Ornstein-Uhlenbeck process [1; 2]. In this case the joint d.f. of x_i and x_{i+1} is given by a non-singular bivariate Gaussian distribution. The results in the present paper are stated completely in terms of the d.f.'s $F_2(x, y)$ and $F_1(x)$ defining the stationary Markoff process and will in particular be valid for observations taken at uniformly spaced time intervals from an Ornstein-Uhlenbeck process.

In this section we shall find the distribution of smallest and largest values in samples $x_1, x_2, \cdots, x_n$ drawn from a random process under assumption A and specified by the bivariate d.f. $F_2(x, y)$ and the associated one dimensional marginal d.f. $F_1(x)$. We first prove Theorem I.

THEOREM I. *Under assumption A, the distribution of largest values in samples of size n is given by the* d.f. $G_n^{(1)}(x) = [F_2(x, x)]^{n-1}/[F_1(x)]^{n-2}$.

To prove this result we note that $G_n^{(1)}(x)$, the probability that the largest value in samples of size n is $\leq x$, is given by

$$G_n^{(1)}(x) = \text{Prob}\,(x_1 \leq x, x_2 \leq x, \cdots, x_n \leq x). \tag{3}$$

To evaluate the right-hand side of (3) we proceed as follows:

$$\begin{aligned}(4)\quad &\text{Prob}\,(x_1 \leq x, x_2 \leq x, \cdots, x_n \leq x) = \\ &\text{Prob}\,(x_1 \leq x, x_2 \leq x, \cdots, x_{n-1} \leq x)\,\text{Prob}\,(x_n \leq x \mid x_1 \leq x, \cdots, x_{n-1} \leq x).\end{aligned}$$

But under assumption A, (4) becomes

$$\begin{aligned}(5)\qquad &\text{Prob}\,(x_1 \leq x, x_2 \leq x, \cdots, x_n \leq x) = \\ &\text{Prob}\,(x_1 \leq x, x_2 \leq x, \cdots, x_{n-1} \leq x)\,\text{Prob}\,(x_n \leq x \mid x_{n-1} \leq x)\end{aligned}$$

or

$$G_n^{(1)}(x) = G_{n-1}^{(1)}(x)\,\text{Prob}\,(x_n \leq x \mid x_{n-1} \leq x). \tag{5$'$}$$

But according to assumption A, and (1) and (2)

$$\begin{aligned}(6)\quad \text{Prob}\,(x_n \leq x \mid x_{n-1} \leq x) &= \text{Prob}\,(x_{n-1} \leq x, x_n \leq x)/\text{Prob}\,(x_{n-1} \leq x) \\ &= F_2(x, x)/F_1(x).\end{aligned}$$

Therefore

$$\begin{aligned}(7)\qquad G_n^{(1)}(x) &= G_{n-1}^{(1)}(x)\,F_2(x, x)/F_1(x) \\ &= G_1^{(1)}(x)\,(F_2(x, x))^{n-1}/(F_1(x))^{n-1} \\ &= (F_2(x, x))^{n-1}/(F_1(x))^{n-2}.\end{aligned}$$

This proves Theorem I.

For $n = 1, 2$, and 3 respectively one gets

$$G_1^{(1)}(x) = F_1(x), \quad G_2^{(1)}(x) = F_2(x, x), \quad G_3^{(1)}(x) = (F_2(x, x))^2/F_1(x). \tag{8}$$

THEOREM II. *Under assumption A, the distribution of smallest values in samples of size n is given by the* d.f.

$$H_n^{(1)}(x) = 1 - \frac{[1 - 2F_1(x) + F_2(x, x)]^{n-1}}{[1 - F_1(x)]^{n-2}}. \tag{9}$$

To prove this result we first note that $H_n^{(1)}(x)$, the probability that the smallest value in samples of size n be $\leq x$ is given by,

$$1 - \text{Prob}\,(x_1 > x, x_2 > x, \cdots, x_n > x).$$

To evaluate $H_n^{(1)}(x)$ we proceed as follows:

$$\begin{aligned}\text{Prob}\,(x_1 > x, x_2 > x, \cdots, x_n > x) = \\ \text{Prob}\,(x_1 > x, x_2 > x, \cdots, x_{n-1} > x)\,\text{Prob}\,(x_n > x \mid x_1 > x, \cdots, x_{n-1} > x).\end{aligned} \tag{10}$$

But under assumption A, (10) becomes

$$\begin{aligned}\text{Prob}\,(x_1 > x, x_2 > x, \cdots, x_n > x) = \\ \text{Prob}\,(x_1 > x, x_2 > x, \cdots, x_{n-1} > x)\cdot\text{Prob}\,(x_n > x \mid x_{n-1} > x).\end{aligned} \tag{11}$$

But

$$\text{Prob}\,(x_n > x \mid x_{n-1} > x) = \text{Prob}\,(x_{n-1} > x, x_n > x)/\text{Prob}\,(x_{n-1} > x). \tag{12}$$

To evaluate Prob $(x_{n-1} > x, x_n > x)$ we note that

$$\begin{aligned}\text{Prob}\,(x_{n-1} > x, x_n > x) + \text{Prob}\,(x_{n-1} \leq x, x_n > x) \\ + \text{Prob}\,(x_{n-1} > x, x_n \leq x) + \text{Prob}\,(x_{n-1} \leq x, x_n \leq x) = 1.\end{aligned} \tag{13}$$

Also

$$\begin{aligned}\text{Prob}\,(x_{n-1} \leq x, x_n > x) + \text{Prob}\,(x_{n-1} \leq x, x_n \leq x) \\ = \text{Prob}\,(x_{n-1} \leq x),\end{aligned} \tag{14}$$

and

$$\begin{aligned}\text{Prob}\,(x_{n-1} > x, x_n \leq x) + \text{Prob}\,(x_{n-1} \leq x, x_n \leq x) \\ = \text{Prob}\,(x_n \leq x).\end{aligned} \tag{15}$$

Recalling that

$$F_2(x, x) = \text{Prob}\,(x_{n-1} \leq x, x_n \leq x) \tag{16}$$

and

$$F_1(x) = \text{Prob}\,(x_{n-1} \leq x) = \text{Prob}\,(x_n \leq x) \tag{17}$$

we get

$$\text{(18)} \qquad \text{Prob}\,(x_{n-1} > x, x_n > x) = 1 - 2F_1(x) + F_2(x, x).$$

Therefore (10) becomes

$$\text{(19)} \quad \text{Prob}\,(x_1 > x, x_2 > x, \cdots, x_{n-1} > x, x_n > x) =$$
$$\text{Prob}\,(x_1 > x, x_2 > x, \cdots, x_{n-1} > x)[1 - 2F_1(x) + F_2(x, x)]/(1 - F_1(x)).$$

Applying the recursion formula (19) successively we obtain

$$\begin{aligned}\text{(20)} \quad &\text{Prob}\,(x_1 > x, x_2 > x, \cdots, x_n > x) = \\ &\text{Prob}\,(x_1 > x)[1 - 2F_1(x) + F_2(x, x)]^{n-1}/[1 - F_1(x)]^{n-1} \\ &= [1 - 2\,F_1(x) + F_2(x, x)]^{n-1}/[1 - F_1(x)]^{n-2}.\end{aligned}$$

Therefore $H_n^{(1)}(x)$, the probability that the smallest value in samples of size n is $\leq x$, is given by:

$$\text{(21)} \qquad H_n^{(1)}(x) = 1 - \frac{[1 - 2F_1(x) + F_2(x, x)]^{n-1}}{[1 - F_1(x)]^{n-2}}.$$

This completes the proof of Theorem II.

In particular for $n = 1, 2$, and 3 respectively the d.f.'s of the smallest value in samples of size n are given by:

$$\text{(22)} \qquad \begin{aligned} &H_1^{(1)}(x) = F_1(x), \qquad H_2^{(1)}(x) = 2F_1(x) - F_2(x, x), \\ &H_3^{(1)}(x) = 1 - \frac{[1 - 2F_1(x) + F_2(x, x)]^2}{1 - F_1(x)}.\end{aligned}$$

3. Distribution of the second largest and second smallest values in samples of size n drawn at equally spaced time intervals from a stationary Markoff process. Under assumption A of Section II we can state the following theorem.

THEOREM III. *Under assumption A the distribution of second largest values in samples of size n, $n \geq 2$, is given by the* d.f. $G^{(2)}(x)$,

$$\begin{aligned} G_n^{(2)}(x) = &[F_2(x, x)]^{n-1}/[F_1(x)]^{n-2} \\ &+ 2[F_2(x, x)]^{n-2}\{F_1(x) - F_2(x, x)\}/[F_1(x)]^{n-2} \\ &+ (n - 2)\,[F_2(x, x)]^{n-3}\,\{F_1(x) - F_2(x, x)\}^2/[F_1(x)]^{n-3}(1 - F_1(x)).\end{aligned}$$

To prove this result we first note that $G_n^{(2)}(x)$, the probability that the second largest value is $\leq x$, is given by

$$\text{(23)} \qquad \begin{aligned} G_n^{(2)}(x) = &\text{Prob}\,(x_1 \leq x, x_2 \leq x, \cdots, x_n \leq x) \\ &+ \text{Prob}\,(x_1 > x, x_2 \leq x, x_3 \leq x, \cdots, x_n \leq x) \\ &+ \text{Prob}\,(x_1 \leq x, x_2 > x, x_3 \leq x, x_4 \leq x, \cdots, x_n \leq x) + \cdots \\ &+ \text{Prob}\,(x_1 \leq x, x_2 \leq x, \cdots, x_{n-2} \leq x, x_{n-1} > x, x_n \leq x) \\ &+ \text{Prob}\,(x_1 \leq x, x_2 \leq x, \cdots, x_{n-1} \leq x, x_n > x).\end{aligned}$$

According to Theorem I

$$\text{(24)} \qquad \text{Prob}\,(x_1 \le x, x_2 \le x, \cdots, x_n \le x) = [F_2(x, x)]^{n-1}/[F_1(x)]^{n-2}.$$

It can readily be shown that

$$\text{(25)} \qquad \begin{aligned} &\text{Prob}\,(x_1 > x, x_2 \le x, x_3 \le x, \cdots, x_n \le x) \\ &\quad = \text{Prob}\,(x_1 \le x, x_2 \le x, \cdots, x_{n-1} \le x, x_n > x) \\ &\quad = [F_2(x, x)]^{n-2}\,\{F_1(x) - F_2(x, x)\}/[F_1(x)]^{n-2}. \end{aligned}$$

It can also be shown that each of the remaining $(n - 2)$ terms on the right-hand side of (23) is equal to

$$\text{(26)} \qquad [F_2(x, x)]^{n-3}\,\{F_1(x) - F_2(x, x)\}^2/[F_1(x)]^{n-3}(1 - F_1(x)).$$

Combining (23), (24), (25), and (26) we get the desired result in Theorem III, i.e.,

$$\text{(27)} \qquad \begin{aligned} G_n^{(2)}(x) &= [F_2(x, x)]^{n-1}/[F_1(x)]^{n-2} \\ &\quad + 2[F_2(x, x)]^{n-2}\,\{F_1(x) - F_2(x, x)\}/[F_1(x)]^{n-2} \\ &\quad + (n - 2)[F_2(x, x)]^{n-3}\,\{F_1(x) - F_2(x, x)\}^2/[F_1(x)]^{n-3}(1 - F_1(x)). \end{aligned}$$

In a similar way one can prove Theorem IV.

THEOREM IV. *Under assumption A, the distribution of second smallest values in samples of size n, $n \ge 2$, is given by the* d.f. $H_n^{(2)}(x)$.

$$\text{(28)} \qquad \begin{aligned} H_n^{(2)}(x) &= 1 - \frac{[1 - 2F_1(x) + F_2(x, x)]^{n-1}}{[1 - F_1(x)]^{n-2}} \\ &\quad - 2\,\frac{[1 - 2F_1(x) + F_2(x, x)]^{n-2}}{[1 - F_1(x)]^{n-2}}\,\{F_1(x) - F_2(x, x)\} \\ &\quad - (n - 2)\,\frac{[1 - 2F_1(x) + F_2(x, x)]^{n-3}}{[1 - F_1(x)]^{n-3}}\,\frac{\{F_1(x) - F_2(x, x)\}^2}{F_1(x)}. \end{aligned}$$

REFERENCES

[1] J. L. DOOB, "The brownian movement and stochastic equations," *Annals of Mathematics*, Vol. 43 (1942), pp. 351.

[2] M. C. WANG AND G. E. UHLENBECK, "On the theory of the brownian motion II," *Reviews of Modern Physics*, Vol. 17 (1945), p. 323.

NOTES

This section is devoted to brief research and expository articles and other short items.

NOTE ON THE CONSISTENCY OF THE MAXIMUM LIKELIHOOD ESTIMATE[1]

By Abraham Wald

Columbia University

1. Introduction. The problem of consistency of the maximum likelihood estimate has been treated in the literature by several authors (see, for example, Doob [1][2] and Cramér [2][3]). The purpose of this note is to give another proof of the consistency of the maximum likelihood estimate which may be of interest because of its relative simplicity and because of the easy verifiability of the underlying assumptions. The present proof has some common features with that given by Doob, insofar that both proofs make no differentiability assumptions (thus, not even the existence of the likelihood equation is postulated) and both are based on the strong law of large numbers and an inequality involving the log of a random variable. The assumptions in the present note are stronger in some respects than those made by Doob, but also the results obtained here are stronger. For the sake of simplicity, the author did not attempt to give the most general results or to weaken the underlying assumptions as much as possible. Remarks on possible generalizations are made in Section 4.

Let $X_1, X_2, \cdots$, etc. be independently and identically distributed chance variables. The most frequently considered case in the literature is that where the common distribution is known, except for the values of a finite number of

---

[1] The author wishes to thank J. L. Doob for several comments and suggestions he made in connection with this note.

[2] According to a communication from Doob, his Theorem 4 is incorrect, but is correct if the class of almost everywhere continuous functions in that theorem is replaced by a suitable class C of functions. The class C can be any one of a variety of classes; for example, the class of bounded almost everywhere continuous functions, or the larger class of almost everywhere continuous functions each of which is less than or equal in modulus to any one of a prescribed sequence of functions with finite expectations. His Theorem 5 on the consistency of the maximum likelihood is then dependent on the class C used in Theorem 4.

[3] The proof given by Cramér [2], pp. 500–504, establishes the consistency of some root of the likelihood equation but not necessarily that of the maximum likelihood estimate when the likelihood equation has several roots. Recently, Huzurbazar [3] showed that under certain regularity conditions the likelihood equation has at most one consistent solution and that the likelihood function has a relative maximum for such a solution. Since there may be several solutions for which the likelihood function has relative maxima, Cramér's and Huzurbazar's results taken together still do not imply that a solution of the likelihood equation which makes the likelihood function an absolute maximum is necessarily consistent.

parameters, $\theta^1, \theta^2, \cdots, \theta^k$. In this note we shall treat the parametric case. For any parameter point $\theta = (\theta^1, \cdots, \theta^k)$, let $F(x, \theta)$ denote the corresponding cumulative distribution function of X_i; i.e., $F(x, \theta) = \text{prob. } \{X_i < x\}$. The totality Ω of all possible parameter points is called the parameter space. Thus, the parameter space Ω is a subset of the k-dimensional Cartesian space.

It is assumed in this note that for any θ, the cumulative distribution function $F(x, \theta)$ admits an elementary probability law $f(x, \theta)$. If $F(x, \theta)$ is absolutely continuous, $f(x, \theta)$ denotes the density at x. If $F(x, \theta)$ is discrete, $f(x, \theta)$ is equal to the probability that $X_i = x$.

Throughout this note the following assumptions will be made.

ASSUMPTION 1. *$F(x, \theta)$ is either discrete for all θ or is absolutely continuous for all θ.*

Before formulating the next assumption, we shall introduce the following notations: for any θ and for any positive value ρ let $f(x, \theta, \rho)$ be the supremum of $f(x, \theta')$ with respect to θ' when $| \theta - \theta' | \leqq \rho$. For any positive r, let $\varphi(x, r)$ be the supremum of $f(x, \theta)$ with respect to θ when $| \theta | > r$. Furthermore, let $f^*(x, \theta, \rho) = f(x, \theta, \rho)$ when $f(x, \theta, \rho) > 1$, and $=1$ otherwise. Similarly, let $\varphi^*(x, r) = \varphi(x, r)$ when $\varphi(x, r) > 1$, and $=1$ otherwise.

ASSUMPTION 2. *For sufficiently small ρ and for sufficiently larger r the expected values $\int_{-\infty}^{\infty} \log f^*(x, \theta, \rho)\, dF(x, \theta_0)$ and $\int_{-\infty}^{\infty} \log \varphi^*(x, r)\, dF(x, \theta_0)$ are finite where θ_0 denotes the true parameter point.*[4]

ASSUMPTION 3. *If $\lim_{i=\infty} \theta_i = \theta$, then $\lim_{i=\infty} f(x, \theta_i) = f(x, \theta)$ for all x except perhaps on a set which may depend on the limit point θ (but not on the sequence θ_i) and whose probability measure is zero according to the probability distribution corresponding to the true parameter point θ_0.*

ASSUMPTION 4. *If θ_1 is a parameter point different from the true parameter point θ_0, then $F(x, \theta_1) \neq F(x, \theta_0)$ for at least one value of x.*

ASSUMPTION 5. *If $\lim_{i=\infty} | \theta_i | = \infty$, then $\lim_{i=\infty} f(x, \theta_i) = 0$ for any x except perhaps on a fixed set (independent of the sequence θ_i) whose probability is zero according to the true parameter point θ_0.*

ASSUMPTION 6. *For the true parameter point θ_0 we have*

$$\int_{-\infty}^{\infty} | \log f(x, \theta_0) |\, dF(x, \theta_0) < \infty.$$

ASSUMPTION 7. *The parameter space Ω is a closed subset of the k-dimensional Cartesian space.*

ASSUMPTION 8. *$f(x, \theta, \rho)$ is a measurable function of x for any θ and ρ.*

It is of interest to note that if we forbid the dependence of the exceptional set on θ in Assumption 3, Assumption 8 is a consequence of Assumption 3, as can easily be verified.

[4] The measurability of the functions $f^*(x, \theta, \rho)$ and $\varphi^*(x, r)$ for any θ, ρ and r follows easily from Assumption 8.

In the discrete case, Assumption 8 is unnecessary. In fact, we may replace $f(x, \theta, \rho)$ everywhere by $\bar{f}(x, \theta, \rho)$ where $\bar{f}(x, \theta, \rho) = f(x, \theta, \rho)$ when $f(x, \theta_0) > 0$, and $\bar{f}(x, \theta, \rho) = 1$ when $f(x, \theta_0) = 0$. Here θ_0 denotes the true parameter point. Since $f(x, \theta_0) > 0$ only for countably many values of $x, \bar{f}(x, \theta, \rho)$ is obviously a measurable function of x.

In the absolutely continuous case, $F(x, \theta)$ does not determine $f(x, \theta)$ uniquely. If Assumptions 3, 5 and 8 hold for one choice of $f(x, \theta)$, they do not necessarily hold for another choice of $f(x, \theta)$. This is in a way undesirable, but assumptions of such nature are unavoidable if we want to insure the consistency of the maximum likelihood estimate. It is, however, possible to formulate assumptions which remain valid for all possible choices of $f(x, \theta)$ and which insure the consistency of the maximum likelihood estimate for a particular choice of $f(x, \theta)$. In this connection the following remark due to Doob is of interest. Let Assumptions 3′ and 5′ be the same as 3 and 5, respectively, except that the exceptional set is permitted to depend on the sequence θ_i. If 3′ and 5′ hold for one choice of $f(x, \theta)$, they also hold for any other choice. Doob has shown that Assumptions 3′ and 5′ insure the existence of a choice of $f(x, \theta)$ for which Assumptions 3, 5 and 8 hold. Thus, one may say that Assumptions 3′ and 5′ are the essential ones and the stronger assumptions 3, 5 and 8 are needed merely to exclude a "bad" choice of $f(x, \theta)$.

2. Some lemmas. In this section we shall prove some lemmas which will be used in the next section to obtain the main theorems. Let θ_0 be the true parameter point. By the expected value Eu of any chance variable u we shall mean the expected value determined under the assumption that θ_0 is the true parameter point. For any chance variable u, u' will denote the chance variable which is equal to u when $u > 0$ and equal to zero otherwise. Similarly, for any chance variable u, the symbol u'' will be used to denote the chance variable which is equal to u when $u < 0$ and equal to zero otherwise. We shall say that the expected value of u exists if $Eu' < \infty$. If the expected value of u' is finite but that of u'' is not, we shall say that the expected value of u is equal to $-\infty$.

LEMMA 1. *For any $\theta \neq \theta_0$ we have*

$$E \log f(X, \theta) < E \log f(X, \theta_0) \tag{1}$$

where X is a chance variable with the distribution $F(x, \theta_0)$.

PROOF. It follows from Assumption 2 that the expected values in (1) exist. Because of Assumption 6, we have

$$E \mid \log f(X, \theta_0) \mid < \infty. \tag{2}$$

If $E \log f(X, \theta) = -\infty$, Lemma 1 obviously holds. Thus, we shall merely consider the case when $E \log f(X, \theta) > -\infty$. Then

$$E \mid \log f(X, \theta) \mid < \infty. \tag{3}$$

Let $u = \log f(X, \theta) - \log f(X, \theta_0)$.[5] Clearly, $E \mid u \mid < \infty$. It i known that for

any chance variable u which is not equal to a constant (with probability one) and for which $E \mid u \mid < \infty$, we have[6]

$$(4) \qquad Eu < \log Ee^u.$$

Since in our case

$$(5) \qquad Ee^u \leqq 1,$$

and since u differs from zero on a set of positive probability (due to Assumption 4), we obtain from (4)

$$(6) \qquad Eu < 0.$$

Thus, Lemma 1 is proved.

We shall now prove the following lemma.

LEMMA 2. $\lim_{\rho=0} E \log f(X, \theta, \rho) = E \log f(X, \theta)$.

PROOF. Let $f^*(x, \theta, \rho) = f(x, \theta, \rho)$ when $f(x, \theta, \rho) \geqq 1$, and $=1$ otherwise. Similarly, let $f^*(x, \theta) = f(x, \theta)$ when $f(x, \theta) \geqq 1$, and $=1$ otherwise. It follows from Assumption 3 that

$$(7) \qquad \lim_{\rho=0} \log f^*(x, \theta, \rho) = \log f^*(x, \theta)$$

except perhaps on a set whose probability measure is zero. Since $\log f^*(x, \theta, \rho)$ is an increasing function of ρ, it follows from (7) and Assumption 2 that

$$(8) \qquad \lim_{\rho=0} E \log f^*(X, \theta, \rho) = E \log f^*(X, \theta).$$

Let $f^{**}(x, \theta, \rho) = f(x, \theta, \rho)$ when $f(x, \theta, \rho) \leqq 1$, and $=1$ otherwise. Similarly, let $f^{**}(x, \theta) = f(x, \theta)$ when $f(x, \theta) \leqq 1$, and $=1$ otherwise. Clearly,

$$(9) \qquad | \log f^{**}(x, \theta, \rho) | \leqq | \log f^{**}(x, \theta) |$$

and

$$(10) \qquad \lim_{\rho=0} \log f^{**}(x, \theta, \rho) = \log f^{**}(x, \theta)$$

for all x except perhaps on a set whose probability measure is zero. The relation

$$(11) \qquad \lim_{\rho=0} E \log f^{**}(X, \theta, \rho) = E \log f^{**}(X, \theta)$$

follows from (9) and (10) in both cases, when $E \log f^{**}(X, \theta)$ is finite and when $E \log f^{**}(X, \theta) = -\infty$. Lemma 2 is an immediate consequence of (8) and (11).

LEMMA 3. *The equation*

$$(12) \qquad \lim_{r=\infty} E \log \varphi(X, r) = -\infty.$$

holds.

---

[5] It is of no consequence what value is assigned to u when $f(x, \theta)$ or $f(x, \theta_0)$ is zero, since the probability of such an event, because of (3), is zero.

[6] This is a generalization of the inequality between geometric and arithmetic means. See, for example, HARDY, LITTLEWOOD, POLYA, *Inequalities*, Cambridge 1934, p. 137, Theorem 184.

PROOF. It follows from Assumption 5 that

$$\lim_{r=\infty} \log \varphi(x, r) = -\infty, \tag{13}$$

for any x (except perhaps on a set of probability 0). Since according to Assumption 2,

$$E \log \varphi^*(X, r) < \infty, \tag{14}$$

and since $\log \varphi(x, r) - \log \varphi^*(x, r)$ and $\log \varphi^*(x, r)$ are decreasing functions of r, Lemma 3 follows easily from (13).

3. The main theorems. We shall now prove the following theorems.

THEOREM 1. *Let ω be any closed subset of the parameter space Ω which does not contain the true parameter point θ_0. Then*

$$\text{prob.} \left\{ \lim_{n=\infty} \frac{\operatorname{Sup}_{\theta \epsilon \omega} f(X_1, \theta) f(X_2, \theta) \cdots f(X_n, \theta)}{f(X_1, \theta_0) f(X_2, \theta_0) \cdots f(X_n, \theta_0)} = 0 \right\} = 1. \tag{15}$$

PROOF. Let r_0 be a positive number chosen such that

$$E \log \varphi(X, r_0) < E \log f(X, \theta_0). \tag{16}$$

The existence of such a positive number follows from Lemma 3. Let ω_1 be the subset of ω consisting of all points θ of ω for which $| \theta | \leqq r_0$. With each point θ in ω_1 we associate a positive value ρ_θ such that

$$E \log f(X, \theta, \rho_\theta) < E \log f(X, \theta_0). \tag{17}$$

The existence of such a ρ_θ follows from Lemmas 1 and 2. Since the set ω_1 is compact, there exists a finite number of points $\theta_1, \cdots, \theta_h$ in ω_1 such that $S(\theta_1, \rho_{\theta_1}) + \cdots + S(\theta_h, \rho_{\theta_h})$ contains ω_1 as a subset. Here $S(\theta, \rho)$ denotes the sphere with center θ and radius ρ. Clearly,

$$0 \leqq \operatorname{Sup}_{\theta \epsilon \omega} f(x_1, \theta) \cdots f(x_n, \theta) \leqq \sum_{i=1}^{h} f(x_1, \theta_i, \rho_{\theta_i}) \cdots f(x_n, \theta_i, \rho_{\theta_i}) + \varphi(x_1, r_0) \cdots \varphi(x_n, r_0).$$

Hence, Theorem 1 is proved if we can show that

$$\text{prob} \left\{ \lim_{n=\infty} \frac{f(X_1, \theta_i, \rho_{\theta_i}) \cdots f(X_n, \theta_i, \rho_{\theta_i})}{f(X_1, \theta_0) \cdots f(X_n, \theta_0)} = 0 \right\} = 1 \quad (i = 1, \cdots, h) \tag{18}$$

and

$$\text{prob} \left\{ \lim_{n=\infty} \frac{\varphi(X_1, r_0) \cdots \varphi(X_n, r_0)}{f(X_1, \theta_0) \cdots f(X_n, \theta_0)} = 0 \right\} = 1. \tag{19}$$

The above equations can be written as

$$\text{(20)} \quad \text{prob}\left\{\lim_{n=\infty} \sum_{\alpha=1}^{n} [\log f(X_\alpha, \theta_i, \rho_{\theta_i}) - \log f(X_\alpha, \theta_0)] = -\infty\right\} = 1 \qquad (i = 1, \cdots, h)$$

and

$$\text{(21)} \quad \text{prob}\left\{\lim_{n=\infty} \sum_{\alpha=1}^{n} [\log \varphi(X_\alpha, r_0(- \log f(X_\alpha, \theta_0)] = -\infty\right\} = 1.$$

These equations follow immediately from (16), (17) and the strong law of large numbers. This completes the proof of Theorem 1.

THEOREM 2. *Let* $\bar{\theta}_n(x_1, \cdots, x_n)$ *be a function of the observations* $x_1, \cdots, x_n$ *such that*

$$\text{(22)} \quad \frac{f(x_1, \bar{\theta}_n) \cdots f(x_n, \bar{\theta}_n)}{f(x_1, \theta_0) \cdots f(x_n, \theta_0)} \geqq c > 0 \textit{ for all } n \textit{ and for all } x_1, \cdots, x_n.$$

Then

$$\text{(23)} \quad \text{prob}\{\lim_{n=\infty} \bar{\theta}_n = \theta_0\} = 1.$$

PROOF. It is sufficient to prove that for any $\epsilon > 0$ the probability is one that all limit points $\bar{\theta}$ of the sequence $\{\bar{\theta}_n\}$ satisfy the inequality $|\bar{\theta} - \theta_0| \leqq \epsilon$. The event that there exists a limit point $\bar{\theta}$ of the sequence $\{\bar{\theta}_n\}$ such that $|\bar{\theta} - \theta_0| > \epsilon$ implies that $\operatorname{Sup}_{|\theta-\theta_0|\geqq\epsilon} f(x_1, \theta) \cdots f(x_n, \theta) \geqq f(x_1, \bar{\theta}_n) \cdots f(x_n, \bar{\theta}_n)$ for infinitely many n. But then

$$\text{(24)} \quad \frac{\operatorname{Sup}_{|\theta-\theta_0|\geqq\epsilon} f(x_1, \theta) \cdots f(x_n, \theta)}{f(x_1, \theta_0) \cdots f(x_n, \theta_0)} \geqq c > 0$$

for infinitely many n. Since, according to Theorem 1, this is an event with probability zero, we have shown that the probability is one that all limit points $\bar{\theta}$ of $\{\bar{\theta}_n\}$ satisfy the inequality $|\bar{\theta} - \theta_0| \leqq \epsilon$. This completes the proof of Theorem 2.

Since a maximum likelihood estimate $\hat{\theta}_n(x_1, \cdots, x_n)$, if it exists, obviously satisfies (22) with $c = 1$, Theorem 2 establishes the consistency of $\hat{\theta}_n(x_1, \cdots, x_n)$ as an estimate of θ.

4. Remarks on possible generalizations. The method given in this note can be extended to establish the consistency of the maximum likelihood estimates for certain types of dependent chance variables for which the strong law of large numbers remains valid.

The assumption that the parameter space Ω is a subset of a finite dimensional Cartesian space is unnecessarily restrictive. Let Ω be any abstract space. All of

our results can easily be shown to remain valid if Assumptions 3, 5 and 7 are replaced by the following one:

ASSUMPTION 9. *It is possible to introduce a distance $\delta(\theta_1, \theta_2)$ in the space Ω such that the following four conditions hold:*

(i) *The distance $\delta(\theta_1, \theta_2)$ makes Ω to a metric space*

(ii) $\lim_{i=\infty} f(x, \theta_i) = f(x, \theta)$ *if* $\lim_{i=\infty} \theta_i = \theta$ *for any x except perhaps on a set which may depend on θ (but not on the sequence θ_i) and whose probability measure is zero according to the probability distribution corresponding to the true parameter point θ_0.*

(iii) *If θ_0 is a fixed point in Ω and* $\lim_{i=\infty} \delta(\theta_i, \theta_0) = \infty$, *then* $\lim_{i=\infty} f(x, \theta_i) = 0$ *for any x.*

(iv) *Any closed and bounded subset of Ω is compact.*

REFERENCES

[1] J. L. DOOB, "Probability and statistics," *Trans. Amer. Math. Soc.*, Vol. 36 (1934).

[2] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946.

[3] V. S. HUZURBAZAR, "The likelihood equation, consistency and the maxima of the likelihood function," *Annals of Eugenics*, Vol. 14 (1948).

ON WALD'S PROOF OF THE CONSISTENCY OF THE MAXIMUM LIKELIHOOD ESTIMATE

BY J. WOLFOWITZ

Columbia University

This note is written by way of comment on the pretty and ingenious proof of the consistency of the maximum likelihood estimate which is due to Wald and is printed in the present issue of the *Annals*. The notation of this paper of Wald's will henceforth be assumed unless the contrary is specified.

The consistency of the maximum likelihood estimate is a "weak" rather than a "strong" property, in the technical meaning which these words have in the theory of probability, i.e., it is a property of distribution functions rather than of infinite sequences of observations. Prof. Wald actually proves strong convergence, which is more than consistency. His proof uses the strong law of large numbers, and he remarks that his method "can be extended to establish consistency of the maximum likelihood estimates for certain types of dependent chance variables for which the strong law of large numbers remains valid." Below we shall use Wald's lemmas to give a proof of consistency which employs only the weak law of large numbers. Not only does this proof have the advantage of being expeditious, but it can be extended to a larger class of dependent chance variables.

The consistency of the maximum likelihood estimate follows from the following

THEOREM. *Let η and ϵ be given, arbitrarily small, positive numbers. Let $S(\theta_0, \eta)$ be the open sphere with center θ_0 and radius η, and let $\Omega(\eta) = \Omega - S(\theta_0, \eta)$. Let*

Wald's Assumptions 1–8 hold. There exists a number $h(\eta)$, $0 < h < 1$, and another positive number $N(\eta, \epsilon)$ such that, for any $n > N(\eta, \epsilon)$,

$$P_0 \left\{ \frac{\sup\limits_{\theta \epsilon \Omega(\eta)} \prod\limits_{i=1}^{n} f(X_i, \theta)}{\prod\limits_{1}^{n} f(X_i, \theta_0)} > h^n \right\} < \epsilon$$

where P_0 is the probability of the relation in braces according to $f(x, \theta_0)$.

PROOF: Proceed exactly as in the proof of Wald's Theorem 1 and obtain r_0, $\rho_{\theta 1}, \cdots, \rho_{\theta_h}$, so that the set theoretic sum of the open spheres $S(\theta_i, \rho_{\theta_i})$, $i = 1, 2, \cdots, h$, covers the compact set which is the intersection of $\Omega(\eta)$ with the sphere $|\theta| \leq r_0$. Define $T(\theta_i)$, $i = 1, \cdots, h + 1$, as follows:

$$-2T(\theta_i) = E \log f(X, \theta_i, \rho_{\theta_i}) - E \log f(X, \theta_0)$$

$(i = 1, \cdots, h)$

$$-2T(\theta_{h+1}) = E \log \varphi(X, r_0) - E \log f(X, \theta_0).$$

If any of the right members above are infinite let $T(\theta_i)$ be one, say. Thus all $T(\theta_i)$ are positive. Applying the weak law of large numbers we have that, for any i such that $1 \leq i \leq h + 1$, there exists a positive number N_i such that, when $n > N_i$,

$$P_0 \left\{ \frac{\prod\limits_{1}^{n} f(X_i, \theta_i, \rho_{\theta_i})}{\prod\limits_{1}^{n} f(X_i, \theta_0)} > \exp(-nT(\theta_i)) \right\} > \frac{\epsilon}{h+1}$$

$(i = 1, \cdots, h)$

$$P_0 \left\{ \frac{\prod\limits_{1}^{n} \varphi(X_i, r_0)}{\prod\limits_{1}^{n} f(X_i, \theta_0)} > \exp(-nT(\theta_{h+1})) \right\} > \frac{\epsilon}{h+1}.$$

From this the theorem follows immediately, with

$$N(\eta, \epsilon) = \max_i N_i$$

$$h(\eta) = \max_i \exp\{-T(\theta_i)\}.$$

The author is obliged to Prof. Wald for his kindness in making his paper available to the author.

A NOTE ON RANDOM WALK

BY HERBERT T. DAVID

The Johns Hopkins University Institute for Cooperative Research

A random walk is defined as a series of discrete steps along the real line, here denoted by I. Each step is represented by the chance variable X, with sectionally continuous density function $f(x)$. The walk begins at any point a of I, and continues until a step carries us outside some subregion Ω of I. In this note, Ω is taken as a finite interval with upper bound D and lower bound $D - y$. The chance variables N and Z are, respectively, the number of steps required to end the walk, and the endpoint of the walk. The range of Z always excludes Ω.

Below, we define $x = D - a$, and consider $E(N)$ as a function $G(x, y)$ of x and y. Under specified conditions, a differential equation (32) is derived, relating $G(0, y)$ and $G(x, y)$.

Let

$$\psi_1(t) = f(t - a) \tag{1}$$

$$\psi_n(t) = \int \cdots (n-1) \cdots \int \prod_{i=1}^{n-1} f(g_i) \, f\left(t - a - \sum_{j=1}^{n-1} g_j\right) dg_1 \cdots dg_{n-1}; \qquad n > 1 \tag{2}$$

where

$$\left[a + \sum_{j=1}^{i} g_j\right] \epsilon \Omega, \quad \text{for } i: 1, 2, \cdots, n-1.$$

Then

$$P\{Z \epsilon w_1, N = n\} = \int_{w_1} \psi_n(t)\, dt \qquad \text{for} \quad w_1 \epsilon \bar{\Omega}$$

$$P\{Z \epsilon w_2, N = n\} = 0 \qquad \text{for} \quad w_2 \epsilon \Omega.$$

Hence

$$P\{N = n\} = \int_{\bar{\Omega}} \psi_n(t)\, dt$$
$$E(N =)\ i \sum_{i=1}^{\infty} \int_{\bar{\Omega}} \psi_i(t)\, dt. \tag{3}$$

The transformation $[h_i = a + \sum_{j=1}^{i} g_j;\ i: 1, \cdots, n-1]$ gives for $\psi_n(t)$ the more convenient expression

$$\psi_n(t) = \int_{\Omega} \cdots (n-1) \cdots \int_{\Omega} f(h_1 - a) \cdot \prod_{i=2}^{n-1} f(h_i - h_{i-1}) f(t - h_{n-1})\, dh_1 \cdots dh_{n-1}. \tag{4}$$

The n-fold integral $\int_I \psi_n(t)\, dt$ is absolutely convergent, hence may be integrated first with respect to t. This gives, keeping the notation of (4)

$$\int_I \psi_n(t)\, dt = \int_\Omega \psi_{n-1}(h_{n-1})\, dh_{n-1}. \tag{5}$$

Assuming that $E(N)$ remains finite for all considered a and Ω, series (3) may be rearranged, giving : $E(N) = \sum_{i=1}^{\infty} B_i$ where

$$B_i = \sum_{j=i}^{\infty} \int_{\bar{\Omega}} \psi_j(t)\, dt.$$

Now, $B_1 = \sum_{i=1}^{\infty} P\{N = i\} = 1$. Also, using (5) and induction on n, it is readily shown that $B_n = \int_\Omega \psi_{n-1}(t)\, dt$, so that

$$E(N) = 1 + \sum_{i=1}^{\infty} \int_\Omega \psi_i(t)\, dt \tag{6}$$

Define transformations $T_n : [g_i = D - h_i\, , i : 1, \cdots , n - 1;\, g_n = D - t]$. Substituting expressions (1) and (4) in (6), transform the jth term of the summation by T_j. This gives

$$E(N) = 1 + \sum_{n=1}^{\infty} \int_0^y \cdots (n) \cdots \int_0^y f(x - g_1) \prod_{i=1}^{n-1} f(g_i - g_{i+1})\, dg_1 \cdots dg_n \tag{7}$$

where $x = D - a$.

By (7), $E(N)$ is a function of x and y; hence we write $E(N) \equiv G(x, y)$.

Define:

$M(k)$: Max $f(t)$ for $|t| \le k$.

K : Any number satisfying $K \le [1 - \epsilon]/M(K)$.

R : Any region $[-\infty < x < \infty;\, 0 \le y \le K]$.

M : Max $f(t)$.

L : Any number satisfying $L \le [1 - \epsilon]/M$.

R' : Any region $[-\infty < x < \infty;\, 0 \le y \le L]$.

In the ensuing argument, we shall assume that

$$(x, y) \in R. \tag{8}$$

This condition restricts certain one-dimensional and two-dimensional variables to regions over which some infinite series are uniformly convergent with respect to these variables. Uniform convergence is required to validate term-by-term differentiations and integrations, and to establish the continuity in one or two variables of certain functions represented by series.

Arguments dealing with the solution of integral equations (17), (20) and (25) are valid only under the more restrictive condition

(9) $$(x, y) \in R'$$

this being the general sufficiency condition for the existence of solutions. However, (17) and (20) enter the argument with respect only to the derivation of equation (21) which could have been derived, though in a more cumbersome manner, by a term by term comparison of the series expressions for $[\lambda_{01}(x, y)]$ $[G(y, y)]$ and for $[G_{01}(x, y)]$ $[\lambda(y, y)]$, this latter approach being valid under (8). Similarly, (25) is used only in obtaining (27), which could have been obtained by a direct manipulation of the series expression for $G(x, y)$, this approach also being valid under (8). Hence, all subsequent derivations hold, as long as $(x, y) \in R$

By (8), we may interchange summation and integration with respect to g_1 in (7). This gives

(10) $$G(x, y) = 1 + \int_0^y f(x - g)G(g, y)\, dg.$$

(11) Assume that $f(t)$ has a continuous derivative everywhere

Then $f(t)$ is continuous and $G(x, y)$ is continuous by (7) and (8). Hence

(12) $f(x - g)G(g, y)$ and $d/dx\, f(x - g)G(g, y)$ are continuous in (x, g)

(13) $f(x - g)G(g, y)$ is continuous in (g, y).

Let $G_{ij}(x, y)$ denote

$$\frac{d^i}{dx^i}\frac{d^j}{dy^j} G(x, y).$$

Then, by (12), we may differentiate (10) with respect to x, and, since $f_{10}(x - g) = -f_{01}(x - g)$, an integration by parts yields

(14) $$G_{10}(x, y) = f(x)G(0, y) - f(x - y)G(y, y) + \int_0^y f(x - g)G_{10}(g, y)\, dg.$$

Further, under (8), $G_{01}(x, y)$ may be obtained by differentiating (7) term by term, and is continuous in (x, y). Hence, $f(x - g)G_{01}(g, y)$ is continuous in (g, y), and we may differentiate (10) with respect to y, giving

(15) $$G_{01}(x, y) = f(x - y)G(y, y) + \int_0^y f(x - g)G_{01}(g, y)\, dg.$$

Adding (14) to (15), dividing by $G(0, y)$ which is always greater or equal to 1, and letting

(16) $$\lambda(x, y) = [G_{10}(x, y) + G_{01}(x, y)]/G(0, y)$$

we obtain

(17) $$\lambda(x, y) = f(x) + \int_0^y f(x - g)\lambda(g, y)\, dg.$$

Under (9), (17) defines a function

$$\lambda(x, y) = f(x) + \sum_{n=1}^{\infty} \int_0^y \cdots (n) \cdots \int_0^y f(x - g_1) \cdot \prod_{i=1}^{n-1} f(g_i - g_{i+1})\, f(g_n)\, dg_1 \cdots dg_n. \tag{18}$$

By (8), this function is continuous in (x, y) and may be differentiated term by term with respect to y. Further, $\lambda_{01}(x, y)$ thus gotten is continuous in (x, y), so that $f(x - g)\lambda_{01}(g, y)$ is continuous in (g, y). Hence, (17) may be differentiated with respect to y, giving

$$\lambda_{01}(x, y) = f(x - y)\lambda(y, y) + \int_0^y f(x - g)\lambda_{01}(g, y)\, dg. \tag{19}$$

Since, under (9), the integral equation

$$\alpha(x, y) = f(x - y) + \int_0^y f(x - g)\alpha(g, y)\, dg \tag{20}$$

has a unique continuous solution for every fixed y, (15) and (19) give

$$\frac{\lambda_{01}(x, y)}{\lambda(y, y)} = \frac{G_{01}(x, y)}{G(y, y)}. \tag{21}$$

Hence

$$\frac{\int_0^y \lambda_{01}(x, y)\, dx}{\lambda(y, y)} = \frac{\int_0^y G_{01}(x, y)\, dx}{G(y, y)}$$

and

$$\frac{\dfrac{d}{dy} \int_0^y \lambda(x, y)\, dx}{\lambda(y, y)} = \frac{\dfrac{d}{dy} \int_0^y G(x, y)\, dx}{G(y, y)}. \tag{22}$$

$$\text{Let } f(t) = f(-t). \tag{23}$$

Then it is obvious from the definition that

$$\underset{\cdot}{G}(0, y) = G(y, y). \tag{24}$$

Further, by (15),

$$\frac{G_{01}(x, y)}{G(y, y)} = f(x - y) + \int_0^y f(x - g) \frac{G_{01}(g, y)}{G(y, y)}\, dg \tag{25}$$

so that, under (9), (25) gives for $G_{01}(x, y)/G(y, y)$ the unique expression

$$f(x - y) + \sum_{n=1}^{\infty} \int_0^y \cdots (n) \cdots \int_0^y f(x - g_1) \prod_{i=1}^{n-1} f(g_i - g_{i+1}) f(g_n - y)\, dg_1 \cdots dg_n$$

which, by (23), is equal to

$$f(y-x) + \sum_{n=1}^{\infty} \int_0^y \cdots (n) \cdots \int_0^y f(y-g_n) \prod_{i=1}^{n-1} f(g_{i+1}-g_i) f(g_1 - x)\, dg_1 \cdots dg_n .$$

Since, under (8), we may interchange summation and integration with respect to x, it follows that

$$\int_0^y \frac{G_{01}(x, y)}{G(y, y)}\, dx = \int_0^y f(y-x)\, dx + \sum_{n=1}^{\infty} \int_0^y \cdots (n+1) \cdots \tag{26}$$
$$\cdot \int_0^y f(y-g_n) \prod_{i=1}^{n-1} f(g_{i+1}-g_i) f(g_i - x)\, dg_1 \cdots dg_n\, dx$$

which, by a change of integration indices and a referral to (7), is seen to equal $[G(y, y) - 1]$. (26) thus gives

$$\int_0^y G_{01}(x, y)\, dx = G(y, y)[G(y, y) - 1]. \tag{27}$$

Further, by (16), (24), and (27),

$$\int_0^y \lambda(x, y)\, dx = G(0, y) - 1 \tag{28}$$

so that

$$\frac{d}{dy} \int_0^y \lambda(x, y)\, dx = \frac{d}{dy} G(0, y) \tag{29}$$

while (24) and (27) also yield

$$\frac{d}{dy} \int_0^y G(x, y)\, dx = [G(0, y)]^2. \tag{30}$$

Hence, by (22), (29), and (30),

$$\lambda(y, y) = \frac{d}{d_y} G(0, y)/G(0, y). \tag{31}$$

Finally, substituting (31) in (21), and remembering the definition of λ given in (16), we get, using (24),

$$G(0, y)[G_{11}(x, y) + G_{02}(x, y)] = \frac{d}{dy} G(0, y)[G_{10}(x, y) + 2G_{01}(x, y)]. \tag{32}$$

The conditions under which (32) holds are, in summary, (8), (11), and (23). If $f(t)$ has an expansion

$$f(t) = \sum_{i=0}^{\infty} A_i t^i; \qquad |t| < T \tag{33}$$

it is clear from (7) that

$$G(x, y) = \sum_{i,j=0}^{\infty} B_{ij} x^i y^j \tag{34}$$

for $(x, y) \in S$, where $S : [T_0 \leq x \leq T_1; 0 \leq y \leq T_1 + T_0]$; $T_0 \leq 0$, $T_1 < T$.

Substituting (34) in (32), and equating coefficients of like powers of (x, y), we obtain the recursion formulae

$$(35) \quad \sum_{j+k=n} B_{ij} B_{0k}[j][2k-j+1] = \sum_{j+k=n-1} B_{i+1,j} B_{0k}[i+1][j-k]; \quad i: 0, 1, \cdots.$$

From (10), it is readily verified that $B_{i0} = 0$ for $i \neq 0$, so that equations (35) give solutions for the B_{ij} in terms of the B_{0k}. These solutions are of interest since they show a one-to-one correspondence between the functions $G(0, y)$ and $G(x, y)$, for $(x, y) \epsilon [R \cap S]$.

NUMERICAL INTEGRATION FOR LINEAR SUMS OF EXPONENTIAL FUNCTIONS

By Robert E. Greenwood

The University of Texas and the Institute for Numerical Analysis[1]

1. Introduction. The methods of numerical integration going by the names trapezoidal rule, Simpson's rule, Weddle's rule, and the Newton-Cotes formulae are of the type

$$(1) \qquad \int_{-1}^{1} f(x)\, dx \simeq \sum_{i=0}^{n} \lambda_{in} f(x_{in})$$

where the abscissae $\{x_{in}\}$ are uniformly distributed on a finite interval, chosen as $(-1, 1)$ for convenience,

$$(2) \qquad x_{in} = -1 + \frac{2i}{n}, \qquad i = 0, 1, 2, \cdots, n,$$

and where the set of constants $\{\lambda_{in}\}$ depend on the name of the rule and the value of n but not on the function $f(x)$. Throughout this note all abscissae will be assumed to be uniformly distributed on $(-1, 1)$ unless the contrary is explicitly stated.

Since correspondence relation (1) involves $(n + 1)$ constants $\{\lambda_{in}\}$, it might be possible to choose $(n + 1)$ arbitrary functions $g_j(x)$, $j = 0, 1, 2, \cdots, n$, and require that the set $\{\lambda_{in}\}$ be the solution, if such exists, of the $(n + 1)$ simultaneous linear equations

$$(3) \qquad \int_{-1}^{1} g_j(x)\, dx = \sum_{i=0}^{n} \lambda_{in} g_j(x_{in}), \qquad j = 0, 1, 2, \cdots, n.$$

Indeed, the selection

$$(4) \qquad g_j(x) = x^j, \qquad j = 0, 1, 2, \cdots, n,$$

will give a set of $(n + 1)$ simultaneous equations of form (3) and the solution $\{\lambda_{in}\}$ is the set of Newton-Cotes weights for that value of n. The numerical evaluation

[1] This work was performed with the financial support of the Office of Naval Research of the Navy Department.

of $\{\lambda_{in}\}$ is best accomplished by other and more sophisticated methods, however.[2]

Because of linearity in both the integral and the finite summation, once the constants $\{\lambda_{in}\}$ have been determined for a specific set of functions $\{g_j(x)\}$, correspondence relation (1) is exact for any linear combination of that fundamental set. Thus, for example, for the fundamental set (4), correspondence relation (1) with the appropriate values $\{\lambda_{in}\}$ is exact for all polynomials of degree less than or equal to n.

Although tradition favors the set of functions (4), there is nothing compelling about such a selection. Indeed, two other possible choices might be

$$g_j(x) = e^{jx}, \qquad j = 0, 1, 2, \cdots, n, \tag{5}$$

and

$$g_j(x) = e^{jx}, \qquad j = -m, -m+1, \cdots, 0, 1, \cdots, m-1, m; n = 2m. \tag{6}$$

These choices would seem to be appropriate whenever numerical methods are being applied to exponential growth curves or exponential decay curves.

2. Use of the basic set $g_j(x) = e^{jx}$. If integration relation (1) be made exact for the set $\{e^{jx}\}$, $j = 0, 1, \cdots, n$ with evenly spaced x abscissae, the set (3) of $(n + 1)$ simultaneous linear equations in the unknowns $\{\lambda_{in}\}$, $i = 0, 1, \cdots, n$ is obtained. Call the solution of this system $\{a_{in}\}$, solution values for $n = 1, 2, 3, 4, 5, 6$ are tabulated below.

For the symmetric case where integration relation (1) is made exact for $\{e^{jx}\}$, $j = -m, -m + 1, \cdots, m - 1, m$; $n = 2m$, a similar but different set of linear equations (3) results for the unknowns $\{\lambda_{in}\}$. Call the solution of this system $\{b_{in}\}$. As implied above, only even values of n are used in order to preserve the symmetry, and values of $\{b_{in}\}$ are tabulated below for $n = 2, 4, 6$.

$n = 1,$	$a_{01} = 1.31303\ 5285$	
	$a_{11} = 0.68696\ 4715$	
$n = 2,$	$a_{02} = 0.21805\ 032^+$	$b_{02} = 0.32260\ 623^-$
	$a_{12} = 1.49780\ 742$	$b_{12} = 1.35478\ 755$
	$a_{22} = 0.28414\ 226^-$	$b_{22} = 0.32260\ 623^-$
$n = 3,$	$a_{03} = 0.51324\ 284$	
	$a_{13} = 0.22445\ 055$	
	$a_{23} = 1.08155\ 527$	
	$a_{33} = 0.18075\ 134$	
$n = 4,$	$a_{04} = -0.13716\ 639^+$	$b_{04} = 0.15048\ 171$
	$a_{14} = 1.40098\ 548$	$b_{14} = 0.73243\ 318$

[2] Whittaker and Robinson, *The Calculus of Observations*, 4th Edition, (1946), London, pp. 152–156.

	$a_{24} = -0.30895\ 914$	$b_{24} = 0.23417\ 022$
	$a_{34} = 0.91710\ 903$	$b_{34} = 0.73243\ 318$
	$a_{44} = 0.12803\ 103^-$	$b_{44} = 0.15048\ 171$
$n = 5,$	$a_{05} = 0.68919\ 3$	
	$a_{15} = -1.07644\ 3$	
	$a_{25} = 2.12534\ 6$	
	$a_{35} = -0.63595\ 6$	
	$a_{45} = 0.79933\ 8$	
	$a_{55} = 0.09852\ 18$	
$n = 6,$	$a_{06} = -0.83607$	$b_{06} = 0.09443\ 5$
	$a_{16} = 3.54128$	$b_{16} = 0.53464\ 7$
	$a_{26} = -3.88102$	$b_{26} = 0.01139\ 3$
	$a_{36} = 3.32254$	$b_{36} = 0.71905\ 0$
	$a_{46} = -0.94685$	$b_{46} = 0.01139\ 3$
	$a_{56} = 0.72075$	$b_{56} = 0.53464\ 7$
	$a_{66} = 0.07937\ 5^+$	$b_{66} = 0.09443\ 5$

The computing service of the Institute for Numerical Analysis has supplied the author with most of the coefficients tabulated above.

3. Estimates of the error term. The choices of the coefficients $\{a_{in}\}$ and $\{b_{in}\}$ are such that integration relation (1) is exact whenever

$$f(x) = A_0 + A_1 e^x + \cdots + A_n e^{nx} \quad \text{and} \quad \lambda_{in} = a_{in}, \tag{7}$$

and whenever

$$f(x) = B_{-m}e^{-mx} + B_{-m+1}e^{-(m-1)x} + \cdots + B_0 + \cdots + B_m e^{mx} \text{ and } \lambda_{in} = b_{in}. \tag{8}$$

When $f(x)$ is not of these prescribed forms, the error in using correspondence (1) may be of some importance. By making the transformation

$$u = e^x, \qquad f(x) = f(\log u) = g(u) \tag{9}$$

integration relation (1) becomes

$$\int_{e^{-1}}^{e} g(u)\frac{du}{u} \simeq \sum_{i=0}^{n} \lambda_{in} g(u_{in}) \tag{10}$$

where the $\{u_{in}\}$ are not evenly distributed. By approximating $g(u)$ by its Taylor's series with a remainder term, the following expressions for the error in using correspondence (1) can be obtained:
Using the coefficients $\{a_{in}\}$,

$$\text{Error} \leq \frac{\left(\dfrac{e^2 - 1}{2e}\right)^{n+1}}{(n + 1)!}\left[2 + \sum_{i=0}^{n} |a_{in}|\right]\left[\max_{-1 \leq x \leq 1}\left(e^{-x}\frac{d}{dx}\right)^{n+1} f(x)\right] \tag{11}$$

and, using the coefficients $\{b_{in}\}$,

$$\text{Error} \le \frac{\left(\dfrac{e^2-1}{2e}\right)^{2m+1}}{(2m+1)!}\left[\frac{e^m - e^{-m}}{m} + \sum_{i=0}^{2m}\frac{|b_{i,2m}|}{e^{mx_{i,2m}}}\right] \cdot \left[\max_{-1\le x\le 1}\left(e^{-x}\frac{d}{dx}\right)^{2m+1} e^{mx} f(x)\right]. \tag{12}$$

Neither of these error expressions can be said to be very practical in actual computation, and neither appears suitable for establishing convergence properties of the type

$$\lim_{n\to\infty}\sum_{i=0}^{n}\lambda_{in} f(x_{in}) = \int_{-1}^{1} f(x)\,dx. \tag{13}$$

However, both (11) and (12) reduce to zero when $f(x)$ is of the form prescribed by (7) or (8) respectively.

4. Numerical examples. As illustrative numerical examples, the case $n = 4$ was selected and several typical functions were integrated approximately by the positive power exponential rule, the symmetrical exponential rule and the Newton-Cotes formula,

$$\int_{-1}^{1} f(x)\,dx = \tfrac{1}{45}[7f(-1) + 32f(-\tfrac{1}{2}) + 12f(0) + 32f(\tfrac{1}{2}) + 7f(1)].$$

Values of $\{a_{i4}\}$ and $\{b_{i4}\}$ are given in the tables in part 2. The typical functions used were x^2, e^{2x}, $1/(x+3)$, e^{-x^2}, xe^x, x^6, and $e^{2.2x}$. The following results were obtained:

Function	*Positive Power Expontential*	*Symmetrical Exponential*	*Newton-Cotes*	*8 Decimal Approximation to Exact Value*
x^2	.5703 8827	.6671 8001	.6666 6666	.6666 6667
e^{2x}	3.6268 6044	3.6268 6041	3.6317 3108	3.6268 6041
$1/(x+3)$	.6828 6353	.6931 5792	.6931 7460	.6931 4718
e^{-x^2}	1.4930 1396	1.4857 2754	1.4887 4582	1.4936 4827
xe^x	.7292 4338	.7353 6007	.7361 7480	.7357 5888
x^6	.0270 8487	.3238 5196	.3333 3332	.2857 1429
$e^{2.2x}$	4.0527 7287	4.0530 7585	4.0607 7415	4.0519 1379

From this tabulation, it would appear that the symmetrical exponential method compares favorably with the Newton-Cotes method for such typical functions as $1/(x+3)$, e^{-x^2}, xe^x, x^6, and $e^{2.2x}$. Note that the choice of x^2 or e^{2x} is not really a fair choice when comparing these two methods, since Newton-Cotes is derived so as to give exactness for x^2 and the symmetrical exponential so as to give exactness for e^{2x}.

SMOOTHEST APPROXIMATION FORMULAS

By Arthur Sard[1]

Queens College

Introduction. Consider a process of approximation which operates on a function $x = x(t)$. The error in the process may be thought of as a sum $R + \delta A$, where R is the error that would be present if x were exact and δA is the error due to errors in x. (Precise definitions are given below.) Suppose that one wishes to choose one process A from a class $\mathfrak{A}$ of processes. In some situations it is appropriate to base the choice on R alone[2]; in others it is appropriate to consider δA.

The primary purpose of the present note is to formulate a criterion of smoothest approximation: That A in $\mathfrak{A}$ is smoothest which minimizes the variance of δA. A criterion based on both R and δA is also suggested. (Sections 1 and 2.) Smoothest approximate integration formulas of one type are derived in Section 3.

Progress in the technique of estimating the covariance function of the errors in x will lead to further applications of the criterion of smoothest approximation.

1. Approximation of a functional. Suppose that X is a space of functions $x = x(t)$ each of which is continuous on $a \leqq t \leqq b$. Let $f[x]$ be a functional defined on X; that is, $f[x]$ is a real number defined for each $x \in X$. For example, X might be the space of functions with second derivatives on $[a, b]$ and $f[x]$ might be $x''(u)$, where u is a fixed number in $[a, b]$.

Suppose that $f[x]$ is to be approximated by a Stieltjes integral

$$A = \int_a^b x(t)\, d\alpha(t), \qquad x \in X, \tag{1}$$

where α is a function of bounded variation. The remainder in the approximation of $f[x]$ by A is

$$R = A - f[x].$$

If the approximation (1) operates on $x + \delta x$ instead of x, the result is $A + \delta A = \int_{t=a}^{b} (x + \delta x)\, d\alpha$; and the error in the approximation of $f[x]$ by $A + \delta A$ is $R + \delta A$, where

$$\delta A = \int_a^b \delta x(t)\, d\alpha(t). \tag{2}$$

Consider a class $\mathfrak{A}$ of approximations A, each of the form (1). We shall propose a criterion for characterizing the "smoothest A" in $\mathfrak{A}$, relative to the covariance function of the errors δx.

---

[1] The author gratefully acknowledges financial support received from the Office of Naval Research.

[2] "Best approximate integration formulas; best approximation formulas," *Amer. Jour. of Math.*, Vol. 71 (1949), pp. 80–91.

Assume that $\delta x = \delta x(t)$ is a stochastic process with mean zero[3] and covariance function $\sigma(t, u) = E[\delta x(t)\delta x(u)]$. Then, by (2), δA is a stochastic variable; and[4]

$$
(3)\qquad
\begin{aligned}
E\delta A &= E\int_{t=a}^{b} \delta x\, d\alpha = \int_{t=a}^{b} 0\, d\alpha = 0,\\
E(\delta A)^2 &= v = E\left[\int_a^b \delta x(t)\, d\alpha(t) \int_a^b \delta x(u)\, d\alpha(u)\right] = \int_a^b\int_a^b \sigma(t,u)\, d\alpha(t)\, d\alpha(u).
\end{aligned}
$$

CRITERION. *That A (if any) in $\mathfrak{A}$ is smoothest which minimizes the variance v of δA.*

In particular cases, this criterion (least squares) has been proposed and used by Chebyshev and others. An application to approximate integration is given in section 3 below.

One may extend this discussion to cases in which the approximations A involve derivatives of x.

Remark. The criterion of *best* approximation[2] may be combined with the above criterion of *smoothest* approximation as follows: That A (if any) in $\mathfrak{A}$ is the best compromise which minimizes a specified combination of the variance of δA and the modulus of R. Here it is assumed that the remainders R satisfy the conditions for the existence of the modulus.[2]

2. Approximation of a function. One may extend the preceding discussion to the case in which $y = f[x]$ is an operation to a space of functions $y = y(u)\cdot$ $\bar{a} \leqq u \leqq \bar{b}$; and in which the approximation of $f[x]$ is

$$
A = \int_a^b x(t)\, d_t\, \alpha(t, u), \qquad x \in X,
$$

where, for each u, α is a function of bounded variation in t. Then, for each u, δA has a variance $v(u)$. *Criterion*. That A (if any) in a class of approximations is smoothest which minimizes $v(u)$ for all u; failing such an A, that A (if any) is smoothest which minimizes the integral of $v(u)$, or alternatively, the supremum of $v(u)$, over $\bar{a} \leqq u \leqq \bar{b}$.

3. Smoothest approximate integration formulas in a particular case.[5] Let m and n be fixed integers; $m \geqq 1$, $n \geqq 0$. *Let $\mathfrak{A} = \mathfrak{A}_{m,n}$ be the class of all approximations of*

---

[3] The essential point here is that $E\delta(t) = m(t)$ be known for each t; for given $m(t)$, one could and would replace $x + \delta x$ by $x + \delta x - m$.

[4] We assume here that the integrals in (3) exist and that the inversions of E and $\int d\alpha$ are valid. For this it is sufficient that δx be integrable relative to the product measure $\alpha\omega$ for all functions α corresponding to elements of $\mathfrak{A}$, where ω is the measure in the underlying probability space relative to which E is the operator $\int d\omega$. Cf. J. L. Doob, "Probability in function space," *Bull. Amer. Math. Soc.*, Vol. 53 (1947), especially pp. 26, 27.

[5] The approximate integration formulas of this section are of such a nature that one would expect them to be known. The values of J at the end are probably new.

$$\int_{-m/2}^{m/2} x(t)\,dt = f[x]$$

of the form

$$A = \sum_{i=-m/2}^{m/2} b_i x(i),$$

the $m + 1$ *constants* b_i *being such that* $A = f[x]$ *whenever* $x(t)$ *is a polynomial of degree* n. Throughout this section i is to range over the $m + 1$ values $i = -m/2, -m/2 + 1, \cdots, + m/2$. *Suppose that the errors* $\delta x(i)$ *are independent, with common variance* σ^2, *and with mean zero.* Then $\alpha(t)$ is a step function with jumps b_i at $t = i$; and

$$v = \sigma^2 \sum_i b_i^2 .$$

The smoothest approximation in $\mathfrak{A}_{m,n}$ is the one for which v is a minimum. (The $m + 1$ variables b_i in v are subject to $n + 1$ constraints due to the condition that the approximation be exact for degree n. The set $\mathfrak{A}_{m,n}$ is empty if and only if m is less than the largest even integer contained in n.)

If $n = 0$ or 1, the smoothest formula in $\mathfrak{A}_{m,n}$ is the one for which all the coefficients are equal:

$$b_i = m/(m + 1);$$

in which case

$$v = m^2\sigma^2/(m + 1).$$

If $n = 2$ or 3, the smoothest formula in $\mathfrak{A}_{m,n}$ is characterized by the following relations:

$$b_i = \lambda_0 + i^2\lambda_1 ,$$

$$\lambda_0 = m(2m^2 + 9m - 6)/2(m - 1)(m + 1)(m + 3),$$

$$\lambda_1 = -30m/(m - 1)(m + 1)(m + 2)(m + 3);$$

in which case

$$v/\sigma^2 = \lambda_0 m + \lambda_1 m^3/12.$$

Thus, the smoothest approximation in $\mathfrak{A}_{6,2}$ or in $\mathfrak{A}_{6,3}$ is the following:

$$A = \tfrac{1}{2}[x(-3) + x(3)] + \tfrac{6}{7}[x(-2) + x(2)] + \tfrac{15}{14}[x(-1) + x(1)] + \tfrac{8}{7}\, x(0).$$

By the method of Lagrange's multipliers, one may establish the following relations for the smoothest formula in $\mathfrak{A}_{m,n}$. Here i has the same range of values as before; μ and ν range over $0, 1, \cdots, [n/2]$.

$$b_i = \sum_\mu \lambda_\mu i^{2\mu},$$

$$v/\sigma^2 = \sum_\mu \lambda_\mu c_\mu ,$$

where

$$c_\mu = m^{2\mu+1}/4^\mu(2\mu + 1),$$

and λ_μ are determined by the equations

$$\sum_\mu \lambda_\mu \sum_i i^{2(\mu+\nu)} = c_\nu .$$

The class $\mathfrak{A}_{m,n}$ is such that for each $A \in \mathfrak{A}_{m,n}$ there is a function $k(t)$ with the following property:[2]

$$R = A - f[x] = \int_{-m/2}^{m/2} x^{(n+1)}(t)k(t)\, dt,$$

whenever x is a function with continuous $(n + 1)$th derivative. The quantity

$$J = \int_{-m/2}^{m/2} k^2(t)\, dt$$

is useful in appraising R, since

$$R^2 \leqq J \int_{-m/2}^{m/2} x^{(n+1)}(t)^2\, dt,$$

by Schwarz's inequality.

Values of J for the smoothest formulas are as follows.

$$n = 0 : J = m^2/6(m + 1).$$

$$n = 1 : J = m^2(3m^2 + 2m + 1)/360(m + 1).$$

For $n = 2$ and 3, and $m \leqq 6$, the numerical values of J are as follows.

m	J $(n = 2)$.	J $(n = 3)$.
2	1/1,890	1/9,072
3	11/8,960	13/17,920
4	134/33,075	62,539/13,891,500
5	1,865/150,528	136,223/6,322,176
6	8/245	6,683/82,320

For the method of calculation of J, as well as the transformation of J under a linear transformation of t, the reader may consult the paper[2].

ON THE POWER FUNCTION OF THE "BEST" t-TEST SOLUTION OF THE BEHRENS-FISHER PROBLEM

By John E. Walsh

The Rand Corporation

1. Introduction. The Behrens-Fisher problem is concerned with significance tests for the difference of the means of two normal populations when the ratio of the variances of the populations is unknown. Denote one population by $N(a_1, \sigma_1^2)$ and the other by $N(a_2, \sigma_2^2)$, where the notation $N(a, \sigma^2)$ represents a normal population with mean a and variance σ^2. Let m sample values be drawn from $N(a_1, \sigma_1^2)$ and n sample values from $N(a_2, \sigma_2^2)$ where $m \leq n$. Then Scheffé [1] has shown that certain optimum properties are possessed by a t-test solution he proposed for the Behrens-Fisher problem, in which the numerator of t is based on the difference of the means of the samples while the denominator is based on the square root of a function of the sample values which has a χ^2-distribution with $m - 1$ degrees of freedom. The purpose of this note is to compare the power function of this t-test with the power function of the corresponding most powerful test for the case in which the ratio of variances σ_1^2/σ_2^2 is also known (only one-sided and symmetrical tests are considered). This comparison is made by computing the power efficiency (see section 2 for definition) of Scheffé's test.

It is sufficient to limit power efficiency investigations to one-sided tests. As shown in [2], a symmetrical t-test with significance level 2α has the same power efficiency as the corresponding one-sided t-test with significance level α. Equation (2) of section 2 furnishes an explicit formula whereby approximate power efficiencies can be computed for a wide range of values of α, m, n. Table 1 contains values of (2) for $\alpha = .05, .01$ and several values of m and n.

For the situation considered here, a power efficiency of $100r\%$ has the quantitative interpretation that the given test based on samples of size m and n has approximately the same power function as the corresponding most powerful test based on samples of size rm and rn. Intuitively the power efficiency of a test measures the percentage of available information per observation which is utilized by that test.

2. Power efficiency derivations. The basic notion of the power efficiency of a significance test is given in [2]. For the present case the problem is to determine the value r such that a most powerful test of the same hypothesis (same significance level) based on rm and rn sample values will have approximately the same power function as the given t-test based on m and n sample values (from $N(a_1, \sigma_1^2)$ and $N(a_2, \sigma_2^2)$ respectively). Here the value of σ_1^2/σ_2^2 is assumed to be known. Then the power efficiency of the given t-test equals $100r\%$.

If the ratio of variances σ_1^2/σ_2^2 is known, the most powerful significance test (one-sided and symmetrical) for the difference of means of the two normal populations is a t-test where the numerator of t is based on the difference of the

TABLE 1

Percentage Power Efficiencies for Certain Values of m and n

$\alpha = .05$

$m \backslash n$	4	6	10	15	20	30	50	100	∞
4	79.6	73.5	67.2	63.4	61.4	59.3	57.6	56.2	54.9
6		86.9	82.9	80.2	78.7	77.0	75.5	74.2	72.9
10			92.6	90.9	89.8	88.6	87.3	86.2	85.0
15				95.2	94.4	93.5	92.5	91.5	90.3
20					96.4	95.7	94.9	94.0	92.9
30						97.7	97.1	96.4	95.3
50							98.6	98.1	97.2
100								99.3	98.6
∞									100.0

$\alpha = .01$

$m \backslash n$	6	8	10	15	20	30	50	100	∞
6	74.9	70.2	66.7	61.2	57.9	54.3	51.1	48.6	45.9
8		81.3	78.8	74.7	72.1	69.1	66.3	63.9	61.4
10			85.3	81.9	79.8	77.2	74.7	72.5	69.9
15				90.4	88.9	87.0	85.0	83.1	80.7
20					92.9	91.4	89.8	88.1	85.8
30						95.3	94.1	92.8	90.7
50							97.2	96.3	94.5
100								98.6	97.3
∞									100.0

two sample means while the denominator is based on the square root of a function of the sample values and σ_1^2/σ_2^2 which has a χ^2-distribution with $m + n - 2$ degrees of freedom [1, p. 43]. Thus the problem is that of comparing the power functions of two t-tests.

As stated in section 1, it is sufficient to consider one-sided tests. We find, using a modification of the normal approximation to the power function of a one-sided t-test given in [3], that Scheffé's one-sided t-test for the Behrens-Fisher problem and the corresponding most powerful one-sided test (σ_1^2/σ_2^2 known) have approximately the same power function when r is chosen so that

$$K_\alpha - \delta\sqrt{r}\{1 - K_\alpha^2/2[(m + n)r - 2]\}^{1/2} = K_\alpha - \delta[1 - K_\alpha^2/2(m - 1)]^{1/2},$$

where α is the significance level of the tests, K_α is the value of the standardized normalized deviate exceeded with probability α, and δ is a function of m, n, a_1, a_2, σ_1^2, σ_2^2 and the given hypothetical value of $a_1 - a_2$ being tested. This condition for the approximate equality of the power functions is reasonably accurate for the following cases: $\alpha = .05$, $m \geq 4$; $\alpha = .025$, $m \geq 5$; $\alpha = .01$, $m \geq 6$; $\alpha = .005$, $m \geq 7$. The accuracy of the approximation increases as m increases.

Hence a value of r such that the two power functions are approximately equal is determined by the equation

$$r\{1 - K_\alpha^2/2[(m + n)r - 2]\} = 1 - K_\alpha^2/2(m - 1). \tag{1}$$

Let

$$A = A(m, \alpha) = 1 - K_\alpha^2/2(m - 1).$$

Then solving (1) for the appropriate root yields

$$r = \frac{1}{2(m + n)}\{2 + (m + n)A + K_\alpha^2/2 + \sqrt{[2 + (m + n)A + K_\alpha^2/2]^2 - 8(m + n)A}\}.$$

Thus the power efficiency of Scheffé's one-sided t-test solution to the Behrens-Fisher problem, for the case in which the ratio of the variances is also known, is approximately equal to

$$\frac{50}{(m + n)}\{2 + (m + n)A + K_\alpha^2/2 + \sqrt{[2 + (m + n)A + K_\alpha^2/2]^2 - 8(m + n)A}\}\%$$

for suitable values of α and m.

REFERENCES

[1] HENRY SCHEFFÉ, "On solutions of the Behrens-Fisher problem based on the t-distribution," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 35–44.

[2] JOHN E. WALSH, "Some significance tests for the median which are valid under very general conditions," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 64–81.

[3] N. L. JOHNSON AND B. L. WELCH, "Applications of the non-central t-distribution," *Biometrika*, Vol. 31 (1940), p. 376.

A NOTE ON FISHER'S INEQUALITY FOR BALANCED INCOMPLETE BLOCK DESIGNS

By R. C. Bose

Institute of Statistics, University of North Carolina

1. An experimental design in which v varieties or treatments are arranged in b blocks, is called a *balanced incomplete block design* if

(i) Each block has exactly k treatments $(k < v)$ no treatment occurring twice in the same block.

(ii) Each treatment occurs in exactly r blocks.

(iii) Any two treatments occur together in exactly λ blocks.

It is easy to see that the parameters v, b, r, k, λ of the design satisfy the relations

$$bk = vr \tag{1.0}$$

$$\lambda(v - 1) = r(k - 1). \tag{1.1}$$

Also it is readily seen that

$$r > \lambda \tag{1.2}$$

for otherwise with any given treatment every other treatment would occur in every block. This would make $k = v$, and the design would become a 'randomised block design'.

Fisher (1940), showed that a necessary condition for the existence of a balanced incomplete block design with v treatments and b blocks is

$$b \geqq v. \tag{1.3}$$

It is the object of this note to give a very simple proof of Fisher's inequality.

2. Consider a balanced incomplete block design with parameters

$$v, b, r, k, \lambda \tag{2.0}$$

and let

$$n_{ij} = 1 \text{ or } 0 \tag{2.1}$$

according as the ith treatment does or does not occur in the jth block. Clearly

$$\sum_{j=1}^{b} n_{ij}^2 = r \tag{2.2}$$

$$\sum_{j=1}^{b} n_{ij} n_{i'j} = \lambda \qquad (i \neq i'). \tag{2.3}$$

If possible let $b < v$. Consider the $v \times v$ matrix

$$N = \begin{bmatrix} n_{11} & n_{12} & \cdots & n_{1b} & 0 & \cdots & 0 \\ n_{21} & n_{22} & \cdots & n_{2b} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ n_{v1} & n_{v2} & \cdots & n_{vb} & 0 & \cdots & 0 \end{bmatrix} \tag{2.4}$$

where the last $v - b$ columns of N consist of zeros. It follows from (2.2) and (2.3) that

$$NN' = \begin{bmatrix} r & \lambda & \cdots & \lambda \\ \lambda & r & \cdots & \lambda \\ \cdots & \cdots & \cdots & \cdots \\ \lambda & \lambda & \cdots & r \end{bmatrix} \tag{2.5}$$

where N' denotes the transpose of N.

$$\det(NN') = \{r + \lambda(v-1)\}\,(r-\lambda)^{v-1} \tag{2.6}$$

But $= kr(r-\lambda)^{v-1}$ from (1.1).

$$\det(NN') = \det N \det N' = 0. \tag{2.7}$$

This makes $r = \lambda$, and contradicts (1.2). Hence the assumption $b < v$ is wrong, and we must have

$$b \geq v \tag{2.8}$$

REFERENCES

[1] R. A. Fisher, "An examination of the different possible solutions of a problem in incomplete blocks," *Annals of Eugenics*, London, Vol. 10 (1940), pp. 52–75.

[2] F. Yates, "Incomplete randomised blocks," *Annals of Eugenics*, London, Vol. 7 (1936), pp. 121–140.

ABSTRACTS OF PAPERS

(Presented September 1, 1949 at Boulder at the Twelfth Summer Meeting of the Institute)

1. **Structure of Statistical Elements.** Duane M. Studley, Foundation Research, Colorado Springs, Colorado.

Research in logical semantics and in practical elementation has set forth the proposition that all words and ideas have set form. As a consequence of this universal proposition all notions and conceptions in statistics should be accessible to set-theoretic analysis and interpretation. This paper explains the results of a preliminary analysis performed on statistical notions and conceptions with a view to a proper organization of definitions and conceptions which will, it is hoped, make possible a better and simpler construction of statistics from a system of basic notions.

2. On the Relative Efficiencies of BAN Estimates. LEO KATZ, Michigan State College, East Lansing, Michigan.

J. Neyman, in the Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability, 1949, proved that χ^2 minimum estimates with either of two alternative definitions of χ^2 are efficient, as also are the maximum likelihood estimates. He also raised the question whether some of these estimates were better than others. This paper bears on that question. In making χ^2 minimum estimates, it is often necessary to avoid small frequencies by grouping together at least one tail of the distribution. It is with respect to the parameters of these *modified* distributions that the χ^2 estimates are efficient. Define *relative efficiency* in these circumstances as the ratio of the variance of an efficient estimator in the unmodified case to that of one in the *modified* case. It is shown that, except for a rectangular probability law, the relative efficiency <1 and, further, it decreases as the tail grouping is made wider. Formulae are given for the relative efficiencies of χ^2 minimum estimators for Binomial and Poisson probability laws and some representative values computed to exhibit these effects.

3. Adjustment of an Inverse Matrix Corresponding to Changes in the Elements of a Given Column or a Given Row of the Original Matrix. JACK SHERMAN and WINIFRED J. MORRISON, The Texas Company Research Laboratories, Beacon, New York.

A simple computational procedure is derived for obtaining the elements b'_{ij} of a nth order matrix (B') which is the inverse of (A'), directly from the elements b_{ij} of a matrix (B) which is the inverse of (A), when (A') differs from (A) only in the elements of one column, say the Sth column. The equations which form the basis of the computation are:

$$b'_{Sj} = \frac{b_{Sj}}{\sum_{i=1}^{n} b_{Sr}\, a'_{rS}}, \qquad j = 1, 2, \cdots n.$$

$$b'_{ij} = b_{ij}\, b'_{Sj} \sum_{r=1}^{n} b_{ir}\, a'_{rS}, \qquad \begin{array}{l} i = 1, 2, \cdots S-1, S+1, \cdots n \\ j = 1, 2, \cdots n. \end{array}$$

Analogous equations are derived for the case that A and A' differ in the elements of a given row rather than a column.

4. On the Problem of Optimum Classification. PAUL G. HOEL, University of California at Los Angeles.

Let f_i, $(i = 1, 2, \cdots, k)$, be the probability density function of population i and let p_i be the probability that population i will be sampled. Assume certain differentiability conditions and moment properties. Then, for known parameters, the probability of a correct classification will be maximixed by choosing the region M_i, which corresponds to classifying into population i, as that part of variable space where $p_i f_i \geq p_j f_j$, $(j = 1, 2, \cdots, k)$. If the parameters are unknown, an asymptotically optimum set of estimates will be given by the set that minimizes a certain form in the covariances. Among uncorrelated estimates, maximum likelihood estimates are seen to be asymptotically optimum.

If weight functions, W_{ij}, are introduced and the expected value of the loss is minimized, the same methods of proof show that the region M_i becomes that part of variable space where $\sum_{r=1}^{k} p_r f_r (W_{rj} - W_{ri}) \geq 0$, $(j = 1, 2, \cdots, k)$, and that the criterion for an asymptotically optimum set of estimates is of the same form as the preceding criterion.

5. **Optimal Linear Prediction of Stochastic Processes whose Covariances are Green's Functions.** C. L. Dolph and M. A. Woodbury, University of Michigan, Ann Arbor.

A method of unbiased, minimal variance, linear prediction is developed for problems similar to those of prediction and filtering treated by Wiener. It differs from these in that, the unbiased condition is imposed, only a finite part of the past is employed, and no stationary assumption is used. It is shown that the special stationary case discussed by Cunningham and Hund, "Random Processes in Problems of Air Warfare" (*Supp. Journal Royal Stat. Soc.*, 1946) succeeds because the correlation function, $e^{-\lambda(t-s)}$, well known to that of the process defined by the Langevian equation, is the Green's function of the homogeneous differential equation formed by letting the adjoint differential operator of the Langevian equation operate on the operator of this equation. This relationship is shown to persist for any physically stable linear differential equation driven by "white noise." The well-known equivalence between integral and differential equations is then extended by use of Stieltjes integrals and used to effect the solutions of the integral equations of the first kind which yield the "optimum" linear prediction. The nonstationary example consisting of purely random motion about a mean linear path in the presence of radar type errors is treated in detail.

6. **The Integral of the Gaussian Distribution over the Area Bounded by an Ellipse.** H. H. Germond, RAND Corporation, Santa Monica, California.

This paper describes the preparation of tables from which to obtain the integral of a bivariate Gaussian distribution over the area of an ellipse. The center of the ellipse need not coincide with the mean of the Gaussian distribution, nor need the axes of the ellipse have any special orientation with respect to the Gaussian distribution.

7. **Theorems on Convergency of Compound Distributions with Symmetric Components.** (*By title*) Maria Castellani, University of Kansas City.

The purpose of this paper is to present some results obtained when operations of convulution in R_1 are concerned with a specific family of distributions. The compound distribution $K(x) = F(x) * G(x)$ is here obtained combining any d.f. $F(x)$ with a d.f. $G(x)$ under the restriction of symmetry, i.e., $G(x + h) + G(x - h) = 1$ for any $h > 0$.

A generalization of Cantelli's Inequalities will enable us to write a preliminary theorem on the following upper and lower bounds:

$$F(a - h) - 2\int_h^\infty dG(y) < K(a) < F(a + h) + 2\int_h^\infty dG(y),$$

$$K(a - h) - 2\int_h^\infty dG(y) < F(a) < K(a + h) + 2\int_h^\infty dG(y),$$

where a is any point in R_1 and $h > 0$.

The theorem is derived assuming the Stieltjes Integral,

$$K(a) = \int_{-\infty}^{+\infty} F(a - y)\, dG(y),$$

is taken as a sum of three integrals connected with three convenient intervals $(-\infty, -h)$, $(-h, h)$, $(h, +\infty)$. When the symmetric component of the convolution is a member of a fam-

ily of normal distributions such as $G_\alpha(x) = \frac{2}{\sqrt{\pi}} \int_{-\infty}^{x} e^{-\alpha^2 y^2}\, dy$, where α is an arbitrary parameter, the use of Cantelli's Inequalities give

$$K_\alpha(a-h) - K_\alpha(a) - \frac{2}{\sqrt{\pi}} \int_{dh}^{\infty} e^{-u^2}\, du < F(a) - K_\alpha(a)$$

$$< K_\alpha(a+h) - K_\alpha(a) + \frac{2}{\sqrt{\pi}} \int_{\alpha h}^{\infty} e^{-u^2}\, du,$$

where $K_\alpha(x) = F(x) * G_\alpha(x)$.

The d.f. $K_\alpha(x)$ is a continuous point function in R_1, with a fr. f$\cdot\gamma(x)$ which is everywhere uniformly continuous. For an arbitrarily small $\eta > 0$, a convenient small h and large α may be found which will enable us to prove the following two theorems:

THEOREM 1: Given any d.f. $F(x)$ in R_1, there exists a convenient continuous d.f. $K_\alpha(x)$ which for $\alpha \to \infty$ converges asypmtotically and uniformly almost everywhere to the given d.f. $F(x)$.

THEOREM 2: Given any d.f. $F(x)$ in R_1, there exists in any continuity bordered interval a convenient uniformly convergent series of continuous functions which asymptotically approach the given $F(x)$.

8. Partial Sums of the Negative Binomial in Terms of the Incomplete Beta-Function. (*By title*) JULIUS LIEBLEIN, Statistical Engineering Laboratory, National Bureau of Standards.

In acceptance sampling a certain size sample is taken at random from a lot of items and the lot is accepted if the number of defective items do not exceed a predetermined number characteristic of the sampling plan. The Statistical Engineering Laboratory has been studying the probabilities that a decision to accept or reject can be made before the sample is completely inspected. Such probabilities are found to involve certain sums apparently not previously treated. In this note the author proves a simple identity connecting these sums which greatly facilitates their computation and shows how they may be written in terms of the well-known incomplete beta-function of Karl Pearson, for which extensive tables are available.

9. Large Sample Tests and Confidence Intervals for Mortality Rates. (*By title*) JOHN E. WALSH, RAND Corporation, Santa Monica, California.

In computing mortality rates from insurance data, the unit of measurement used is frequently based on number of policies or amount of insurance rather than on lives. Then the death of one person may result in several units of "death" with respect to the investigation; moreover, the number of units per individual may vary noticeably. Thus the usual large sample methods of obtaining significance tests and confidence intervals for the true value of the mortality rate are not applicable to these situations. If the number of units associated with each person in the ivestigation were known, accurate large sample results could be obtained; however, determination of the number of units associated with each individual would require an extremely large amount of work. This article presents some valid large sample tests and confidence intervals for the mortality rate which do not require much work and are reasonably efficient. The procedure followed consists in first dividing the risks into twenty-six subgroups on the basis of the first letter of the last name of the person insured. Some of the groups are then combined until 10 to 15 subgroups yielding approximately the same number of units are obtained. The fraction consisting of the total number of units paid divided by the total number of units exposed is computed

for each subgroup. Asymptotically the resulting observations represent independent observations from continuous symmetrical populations with common median equal to the true value of the rate of mortality. Tests and confidence intervals for the rate of mortality are obtained by applying the results of the paper "Some Significance Tests for the Median which are Valid Under Very General Conditions" (*Annals of Math. Stat.*, Vol. 20 (1949), pp. 64–81 to these observations.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest

Personal Items

Mr. Fred C. Andrews will be a teaching assistant in the Statistical Laboratory, Department of Mathematics, University of California for the academic year 1949–1950.

Dr. Joseph Berkson has been promoted to the rank of Professor in the University of Minnesota Graduate School and Mayo Foundation. He continues as Chief of the Division of Biometry and Medical Statistics of the Mayo Clinic.

Mr. Colin R. Blyth is now a research assistant at the University of California, Statistical Laboratory, Berkeley.

Mr. Clyde A. Bridger is now Director of the Section of Statistics and State Registrar of Vital Statistics for the Division of Health of Missouri.

Mr. Loren V. Burns, formerly with the MFA Milling Company at Springfield, Missouri, has been made Vice-President and Technical Director of the Spear Mills, Inc., Kansas City 6, Missouri.

Professor Douglas Chapman, who obtained his Ph.D. in statistics at the University of California, Berkeley, has accepted an appointment as Assistant Professor at the University of Washington in the Department of Mathematics and the Laboratory of Statistical Research.

Dr. Andrew Laurence Comrey, who received his doctor's degree from the University of Southern California last June, has accepted an assistant professorship in the Department of Psychology at the University of Illinois.

Dr. Donald A. Darling has been appointed to an instructorship in the Department of Mathematics, University of Michigan.

Dr. Paul M. Densen resigned his position as Chief of the Division of Medical Research Statistics of the Department of Medicine and Surgery of the Veterans Association as of July 1, 1949 to join the staff of the Graduate School of Public Health, University of Pittsburgh, as an Associate Professor of Biostatistics.

Mr. Amron H. Katz has been promoted to the position of Chief Physicist of the Photographic Laboratory, Engineering Division, Air Material Command, Wright Patterson Air Force Base, Dayton, Ohio.

Associate Professor Louis Guttmann, who had been on leave for two years from the Department of Sociology of Cornell University conducting a research program in Israel, was invited to remain in Israel for another year to direct the

activities of the recently founded Israel Institute of Public Opinion Research. He is serving as Chief Consultant.

Mr. Herne Ernest LaFontant who was attending the University of Michigan during the academic year 1948–1949 working on his doctor's degree, has accepted a position as statistician for the B.T.W. Insurance Co. at Birmingham, Alabama.

Assistant Professor Jerome C. R. Li has been promoted to Associate Professor of Mathematics at the Oregon State College, Corvallis, Oregon.

Professor H. B. Mann of Ohio State University has accepted a visiting professorship and research associateship at the Statistical Laboratory at Berkeley, California for the year 1949–1950.

Dr. Gottfried E. Noether has been appointed to an instructorship at New York University.

Dr. G. R. Seth has just returned from a trip to England, Sweden, France and India where he visited statistical institutions.

Assistant Professor Andrew Sobczyk has been promoted to Associate Professor of Mathematics at Boston University.

Dr. Zenon Szatrowski, formerly with the Economics Department of Northwestern University, has accepted an associate professorship in the School of Business Administration, University of Buffalo.

Professor Gerhard Tintner has returned to his teaching and research duties at Iowa State College after spending a year at the Department of Applied Economics at the University of Cambridge, England. He gave a course on Econometrics at the University of Cambridge and during his stay in Europe, he lectured on econometric and statistical subjects in Universities at Bristol, Dublin, Hull, Paris, Manchester and Uppsala.

Dr. A. E. R. Westman, Director of the Department of Chemistry, Ontario Research Foundation, left in September, 1949 for England where he is visiting industrial research laboratories and engaging in studies in the Department of Physical Chemistry, Cambridge University. He plans to return in June, 1950.

Word has just been received here of the formation of the New Zealand Statistical Association. The initial meeting was held in August, 1948 at Victoria University College. The officers are: J. T. Campbell, President; I. D. Dick, Secretary. It is planned to hold one formal meeting a year at first with the hope of increasing this later. The main interest in statistical work in New Zealand has been biological, but there is scope for considerable extension to industrial, educational and economic fields and it is hoped the formation of the Association will assist in this extension.

New Members

The following persons have been elected to Membership in the Institute (June 1, 1949 to August 22, 1949)

Al-Doori, Younis A., Student at the University of California, *1915 Henry Street, Berkeley, California.*

Bieber, Robert A., A.B. (Univ. of Calif.) *S-18 Richmond Terrace, Richmond, California.*

Bula, Clotilde Angelica, Ph.D., (Univ. of Rosario, Argentina) Professor, University of Buenos Aires, *Rioja 3681, Olivos-Pcia. de Buenos Aires, Argentina.*

Dalziel, Edwin R., Ph.D. (Univ., Edinburgh) Assistant Master, Palmerston North Technical School, Palmerston North, New Zealand.

Douglas, James B., Dip. Ed. (Melbourn Univ.) Lecturer in Mathematics, Newcastle Technical College, Tighe's Hill 2N, N.S.W., Australia.

Hartley, Herman O., Ph.D. (Cambridge Univ.) Lecturer in Statistics, Department of Statistics, University College, London, W.C.1, England.

Immel, Eric R., M.A. (Queen's Univ., Kingston, Canada) Teaching Assistant and Graduate Student, Department of Mathematics, University of California at Los Angeles, Los Angeles, California.

Kelly, John P., Senior Technical Engineer, Carbide and Carbon Chemical Corporation, Oak Ridge, Tennessee, *P.O. Box 473, Norris, Tennessee.*

Parel, Cristina P., M.S. (Univ. of Michigan) Instructor, Department of Mathematics, University of the Philippines, Manila, P.I.

Philipson, Carl O., D.Sc. (Univ. of Stockholm) Actuary of Folket-Samarbete, *Yngvevagen 5, Djursholm, Sweden.*

Porter, Robert A., Ph.D. (N.C. State College, Raleigh, N.C.) Senior Mathematician, University of Chicago, *17113 Longfellow Avenue, Homewood, Illinois.*

Rippe, Dayle D., M.A. (Univ. of Nebr.) Student, Teaching Fellow, Department of Mathematics, University of Michigan, *1049 Woburn Court, Willow Run, Michigan.*

Rogers, Robert L., A.B. (Univ. of Calif.) Student at University of California, *Route 2, Box 74, Denio Avenue, Gilroy, California.*

Roy, Samarendra N., M.Sc. (Calcutta Univ.) Head of Department of Statistics, Calcutta University and Assistant Director, Indian Statistical Institute (now on leave) *P.O. Box 168, Chapel Hill, North Carolina.*

Savey, Rosemary, M.B.A. (Univ. of Wisc.) Graduate Assistant and Student, University of Wisconsin, *2513 Norwood Place, Madison 5, Wisconsin.*

REPORT ON THE BOULDER MEETING OF THE INSTITUTE

The Twelfth Summer Meeting of the Institute of Mathematical Statistics was held at the University of Colorado, Boulder, Colorado, Monday, August 29 through Thursday, September 1, 1949. The meeting was held in conjunction with the summer meetings of the American Mathematical Society, the Mathematical Association of America, and the Econometric Society. The meeting was attended by the following 79 members of the Institute:

S. P. Agarwal, R. L. Anderson, T. W. Anderson, V. L. Anderson, K. J. Arnold, E. W. Barankin, C. A. Bennett, Agnes Berger, E. E. Blanche, A. H. Bowker, J. C. Brixey, Jean Bronfenbrenner, J. H. Bushey, H. C. Carver, Herman Chernoff, K. L. Chung, A. G. Clark, E. P. Coleman, E. L. Crow, J. H. Curtiss, W. J. Dixon, J. L. Doob, Aryeh Dvoretzky, H. P. Evans, W. D. Evans, W. T. Federer, William Feller, C. H. Fischer, J. S. Frame, T. C. Fry, H. M. Gehman, H. H. Germond, R. E. Greenwood, H. T. Guard, P. R. Halmos, J. L. Hodges, P. G. Hoel, Harold Hotelling, J. M. Howell, C. C. Hurd, C. A. Hutchinson, Irving Kaplansky, Leo Katz, H. S. Konijn, T. C. Koopmans, G. M. Kuznets, H. D. Larsen, D. H. Leavens, S. B. Littauer, H. B. Mann, Jacob Marschak, F. J. Massye, Dorothy J. Morrow, Jerzy Neyman, M. L. Norden, J. I. Northam, E. G. Olds, R. P. Peterson, G. B. Price, Mina S. Rees, P. R. Rider, F. D. Rigby, Herman Rubin, L. J. Savage, Elizabeth R. Scott,

I. E. Segal, Esther Seiden, Jack Sherman, W. B. Simpson, Milton Sobel, D. M. Studley, B. R. Suydam, A. G. Swanson, James Templeton, R. M. Thrall, J. W. Tukey, Abraham Wald, John Wishart, S. S. Wilks.

The Monday afternoon session was devoted to invited addresses with Professor Leonard J. Savage of the University of Chicago presiding. The attendance was approximately fifty. Professor J. L. Hodges of the University of California presented a paper, *Some Problems in Point Estimation*, and Professor W. T. Federer of Cornell University presented *A Comparison of the Proportionality of Covariance Matrices.*

On Tuesday Morning the Institute, the Mathematical Association of America, and the Econometric Society held a joint symposium on *Mathematical Training for Social Scientists.* Professor Jacob Marschak of the Cowles Commission for Research in Economics presided. The attendance was approximately one hundred fifty. The participating speakers were: Professor R. L. Anderson of North Carolina State College; Professor T. W. Anderson of Columbia University; Professor G. C. Evans of the University of California; Professor F. L. Griffin of Reed College; Professor Harold Gulliksen of Educational Testing Service; Professor William Jaffé of Northwestern University; Professor Harold Hotelling of the University of North Carolina; and Professor G. M. Kuznets of the University of California. At the end of the session the following resolution was adopted by those in attendance at the meeting:

> Members of the Mathematical Association of America, the Institute of Mathematical Statistics, and the Econometric Society assembled in a joint session in Boulder, Colorado, on August 30, 1949, are of the opinion that officers of these societies should study the need for better mathematical training of social scientists, and the ways and means to improve mathematical preparation of social scientists, and that such a study may be most effectively conducted by a joint committee, possibly in co-operation with other interested societies, and in close touch with the Social Science Research Council, the National Research Council, or other national bodies concerned with general education and research. It is suggested that this committee report the results of its deliberations at the next joint meeting of the original participating societies.

The two joint sessions of the Institute and the Econometric Society were devoted to a Symposium on *Statistical Inference in Decision Making.* Professor Jerzy Neyman of the University of California presided on Tuesday afternoon. The attendance was approximately eighty. Professor Aryeh Dvoretzky of Hebrew University, Jerusalem presented *Decision Problems* and Professor Abraham Wald of Columbia University presented *Some Recent Results in the Theory of Statistical Decision Functions.* On Wednesday Morning, under the chairmanship of Professor Wald and an attendance of approximately seventy-five, the following papers were presented: *Remarks on a Rational Selection of a Decision Function* by Professor Herman Chernoff of the Cowles Commission for Research in Economics; *Psychological Probabilities* by Professor Leonard J. Savage; and *Complete Classes of Decision Functions for Some Standard Sequential and Nonsequential Problems* by Milton Sobel of Columbia University.

On Thursday Morning the Institute and the American Mathematical Society held a joint session for contributed papers with Professor P. R. Rider of Washington University presiding. The attendance was approximately seventy-five. The following papers were presented:

1. *Structure of Statistical Elements.*
   Mr. Duane M. Studley, Foundation Research, Colorado Springs.
2. *On the Relative Efficiencies of BAN Estimates.*
   Professor Leo Katz, Michigan State College.
3. *Adjustments of an Inverse Matrix Corresponding to Changes in the Elements of a Given Column or a Given Row of the Original Matrix.*
   Dr. Jack Sherman and Miss Winifred J. Morrison, The Texas Company Research Laboratories, Beacon, New York.
4. *On the Problem of Optimum Classification.*
   Professor Paul G. Hoel, University of California at Los Angeles.
5. *Optimal Linear Prediction of Stochastic Processes whose Covariances are Green's Functions.*
   Professor C. L. Dolph and Dr. M. A. Woodbury, University of Michigan.
6. *The Integral of the Gaussian Distribution over the Area Bounded by an Ellipse.*
   Dr. H. H. Germond, Rand Corporation, Santa Monica, California.
7. *Theorems on Convergency of Compound Distributions with Symmetric Components.* (By title)
   Dr. Maria Castellani, University of Kansas City.
8. *Large Sample Tests and Confidence Intervals for Mortality Rates.* (By title)
   Dr. J. E. Walsh, Rand Corporation, Santa Monica, California.
9. *Partial Sums of the Negative Binomial in Terms of the Incomplete Beta-function.* (By title)
   Dr. Julius Lieblein, National Bureau of Standards.

On Thursday afternoon Professor Jerzy Neyman presented the Second Rietz Memorial Lecture on *Consistent Estimates of the Linear Structural Relation in the General Case of Identifiability.* Professor Harold Hotelling presided and the attendance was approximately fifty. Dr. R. P. Boas, Jr. of Mathematical Reviews presented an invited address *The Representation of Probability Distributions by Charlier Series.*

The Institute sponsored a beer party on Tuesday evening and on Thursday evening a fry was held on Flagstaff Mountain.

HARRIS T. GUARD
Assistant Secretary

THE INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

OFFICERS FOR 1949

President:
J. NEYMAN, University of California, Berkeley

President-Elect:
J. L. DOOB, The University of Illinois, Urbana

Secretary-Treasurer:
C. H. FISCHER, University of Michigan, Ann Arbor

Editor:
S. S. WILKS, Princeton University, Princeton, N. J.

The purpose of the Institute of Mathematical Statistics is to stimulate research in the mathematical theory of statistics and to promote coöperation between the field of pure research and the fields of application.

Membership dues including subscription to the ANNALS OF MATHEMATICAL STATISTICS are $7.00 per year within the Western Hemisphere and $5.00 per year elsewhere. Dues and inquiries regarding membership in the Institute should be sent to the Secretary-Treasurer of the Institute.

MEETINGS OF THE INSTITUTE

ANNUAL MEETING—NEW YORK CITY—December 26-31, 1949

To be held in conjunction with the meetings of the American Statistical Association.

Abstracts must be in the hands of Associate Secretary S. B. Littauer, Department of Industrial Engineering, Columbia University, New York 27, N. Y., not later than November 15.

CHAPEL HILL, N. C., March 17-19, 1950.

Joint Meeting with the Biometric Society. Abstracts of papers in Mathematical Statistics must be in the hands of Professor Herbert Robbins, Department of Mathematical Statistics, Chapel Hill, N. C., not later than February 1. Abstracts of Biometric papers must be in the hands of Professor H. L. Lucas, Department of Experimental Statistics, State College, Raleigh, by February 1.